

NDMA Contaminated Valsartan

David Madigan, PhD

**1. Background**

1. I am Provost and Senior Vice-President for Academic Affairs at Northeastern University. From 2007 to 2020 I was Professor of Statistics at Columbia University in New York City. I was also chair of the Columbia Department of Statistics from 2008 to 2013 and Executive Vice-President of Arts of Sciences and Dean of the Faculty of Arts and Sciences from 2013 to 2018. I received my bachelor's degree in Mathematics from Trinity College Dublin in 1984 and was awarded the College's gold medal. In 1990, I received a Ph.D. in Statistics, also from Trinity College. I have worked in the past for KPMG, SkillSoft, University of Washington, AT&T Labs, and Soliloquy Inc. From 2005 to 2007 I was Professor of Statistics and Dean of Physical and Mathematical Sciences at Rutgers University. Prior to serving as Dean I was Director of the Rutgers University Institute of Biostatistics. I am an elected Fellow of both the Institute of Mathematical Statistics and the American Statistical Association, as well as the American Association for the Advancement of Science, and was the 36th most cited mathematician worldwide from 1995-2005. I was an Institute of Mathematical Statistics Medallion Lecturer in 2009. I served a term as the Editor of *Statistical Science* from 2008 to 2010, the highest impact journal in Statistics.
2. I have published more than 180 technical papers on Bayesian statistics, biostatistics, pharmacovigilance, statistical graphics, Monte Carlo methods, computer-assisted learning, information retrieval, and text mining. Within the last few years, I have consulted for Heron Therapeutics, Shire, Bayer, and Eli Lilly on a variety of issues, many related to drug safety. In the past, I have consulted for Pfizer. I have considerable statistical experience with clinical trials including the design and analysis of pain studies at the University of Washington and the Fred Hutchinson Cancer Research Center, and more generally as a statistical consultant to multiple internal and external clients, particularly while I was director of the Institute of Biostatistics at Rutgers University, and continuing with Shire and Bayer.
3. Drug safety is one of my significant research interests, with a focus on the development and application of statistical methods for pharmacovigilance. I have published my work in *Drug Safety*, *Pharmacoepidemiology and Drug Safety*, *Therapeutic Advances in Drug Safety*, *Epidemiology*, the *American Journal of Epidemiology*, and other journals. I have also served as an investigator in the Mini-Sentinel project. Mini-Sentinel was a pilot project sponsored by the FDA to inform and facilitate development of a fully operational active surveillance system, the Sentinel System, for monitoring the safety of FDA-regulated medical products. In 2010-11, I led the Mini-Sentinel Working Group on case-based methods in active surveillance. In addition, from 2010 to 2013 I was a Principal Investigator for the Observational Medical Outcomes Partnership (OMOP), a public-private partnership between the FDA and the pharmaceutical industry. The partnership conducted a multi-year initiative to research methods that are feasible and useful to analyze existing healthcare databases to identify and evaluate safety and benefit issues of drugs already on the market. The OMOP work now continues in the Observational Health Data Sciences and Informatics (OHDSI) collaborative where I co-

direct the Columbia-based coordinating center. I was a member of the FDA's Drug Safety and Risk Management Advisory Committee (DSaRM) from 2011 to 2014 and then served the FDA as a consultant through 2017. DSaRM advises the FDA Commissioner on risk management, risk communication, and quantitative evaluation of spontaneous reports for drugs for human use and for any other product for which the FDA has regulatory responsibility. From 2010 to 2011 I was a member of a sub-committee of the FDA Science Board charged with reviewing the Center for Drug Evaluation and Research's pharmacovigilance program.

4. Further information concerning my background, training, and experience, including a complete list of my publications, is reflected in my curriculum vitae, a copy of which is attached as Appendix 1. A list of the deposition and trial testimony I have provided in the last four years is attached as Appendix 2.
5. For my services, I am being compensated at the rate of \$800 per hour. My compensation is not contingent on the outcome of this matter. The approach I have taken to this work would have been the same had the defendants hired me to carry out these analyses.

2. Introduction

6. According to the U.S. Department of Health and Human Services, N-nitrosodimethylamine (NDMA), is reasonably anticipated to be a human carcinogen.¹ Similarly, the International Agency for Research on Cancer classifies NDMA as "probably carcinogenic to humans."² The U.S. Food and Drug Administration has indicated that levels of NDMA up to 0.096 µg/day and levels of NDEA up to 0.0265 µg/day are safe.³ More recently Johnson et al.⁴ have suggested that NDMA levels as high as 6.2 µg/day and NDEA levels as high as 2.2 µg/day could be safe. Levels of NDMA in contaminated valsartan tablets range from below the limit of detection to 20.19 µg while levels of NDEA in contaminated valsartan tablets range from below the limit of detection to 1.31 µg.⁵ Other sources show levels as high as 60.2 µg of NDMA⁶ and 5.4 µg of NDEA.⁷

¹ https://www.epa.gov/sites/production/files/2017-10/documents/ndma_fact_sheet_update_9-15-17_508.pdf

² International Agency for Research on Cancer. *Some N-nitroso compounds*. Lyon: IARC, 1978.

³ <https://www.fda.gov/downloads/Drugs/DrugSafety/UCM615703.pdf>

⁴ Johnson, G. E., Dobo, K., Gollapudi, B., Harvey, J., Kenny, J., Kenyon, M., Lynch, A., Minocherhomji, S., Nicolette, J., Thybaud, V., Wheeldon, R., & Zeller, A. (2021). Permitted daily exposure limits for noteworthy N-nitrosamines. *Environmental and Molecular Mutagenesis*.

⁵ Snodin, D. J., & Elder, D. P. (2019). Short commentary on NDMA (N-nitrosodimethylamine) contamination of valsartan products. *Regulatory Toxicology and Pharmacology*, 103, 325-329. Also: <https://www.fda.gov/drugs/drug-safety-and-availability/laboratory-analysis-valsartan-products>

⁶ ZHP Exhibit 118/ SOLCO00028261

⁷ TORRENT-MDL2875-00133890

7. I was asked to perform a statistical analysis to evaluate the strength of association, dose-response and increased risk of the cancers reported in certain dietary and occupational studies which specifically examined exposure to NDMA and/or NDEA. Specifically, I considered the studies considered by Dr. Etminan that estimated NDMA or NDEA effect sizes. For each study, I carefully considered the study design as well the statistical methods that were deployed. These are all observational studies and I have discussed the strengths and limitations of observational studies in several publications.⁸

3. Evidence from Dietary Studies

8. Many dietary studies have considered the potential relationship between NDMA and cancer. Such studies typically rely on dietary questionnaires along with estimated NDMA content for different food types. I considered relevant meta-analyses as well as specific individual studies considering dietary NDMA and its relationship to different cancers.⁹ I also considered a single dietary NDEA study. In each case, I focused primarily on the comparison of the highest NDMA group to the lowest NDMA group with the groups defined by tertiles, quartiles, or quintiles, depending on the study. For each study I compute the mean “lifetime cumulative exposure” (LCE) as the average number of days from birth to study end multiplied by the lower bound of the NDMA (or NDEA) daily level in the highest group. In studies presenting multiple analyses I focused on the maximally adjusted analyses.

Gastric Cancer

9. A 2015 meta-analysis by Song et al.¹⁰ included 11 studies concerning NDMA and gastric cancer and yielded a relative risk estimate of 1.34 and an associated 95% confidence interval of (1.02 , 1.76). In a dose-response analysis, the authors noted a nonlinear trend toward gastric cancer risk with increasing NDMA intake. There was, however, considerable between-study heterogeneity. LCE’s across the component studies ranged from 1,412 µg to 6,607 µg.
10. I note that Loh et al.¹¹ reported a stomach cancer hazard ratio of 1.13 (0.81 , 1.57). Adding Loh et al. to the Song meta-analysis yields an estimate of 1.32 (1.03 , 1.68) and a slightly lower I² of 73.8% (R code in Appendix A).

⁸ See, for example, Madigan, D., Stang, P. E., Berlin, J. A., Schuemie, M., Overhage, J. M., Suchard, M. A., ... & Ryan, P. B. (2014). A systematic statistical approach to evaluating evidence from observational studies. *Annual Review of Statistics and Its Application*, 1, 11-39.

⁹ Gastric, esophageal, pancreatic, lung, colorectal, prostate, bladder

¹⁰ Song, P., Wu, L., & Guan, W. (2015). Dietary nitrates, nitrites, and nitrosamines intake and the risk of gastric cancer: a meta-analysis. *Nutrients*, 7(12), 9872-9895.

¹¹ Loh, Y. H., Jakyszyn, P., Luben, R. N., Mulligan, A. A., Mitrou, P. N., & Khaw, K. T. (2011). N-nitroso compounds and cancer incidence: the European Prospective Investigation into Cancer and Nutrition (EPIC)–Norfolk Study. *The American Journal of Clinical Nutrition*, 93(5), 1053-1061.

Esophageal Cancer

11. A 2016 meta-analysis by Cui et al.¹² found an increased risk of esophageal cancer associated with NDMA, hazard ratio = 1.18, 95% confidence interval (0.98 , 1.41), P=0.07.
12. Cui et al. appear to have not included Rogers et al.¹³ who reported an odds ratio of 1.86, 95% confidence interval (0.87 , 3.95). They also reported a statistically significant odds ratio of 1.82 for a broader group of oral cavity cancers, P=0.12 for trend. I note that a combined analysis¹⁴ of Cui et al. and Rogers et al. yields an estimate of 1.27, 95% confidence interval (0.91 , 1.76), code in Appendix A.

Pancreatic Cancer

13. Zheng et al.¹⁵ reported an odds ratio of 1.03, 95% confidence interval (0.78 , 1.37) for NDMA overall but 1.93, 95% confidence interval (1.42 , 2.61), for NDMA from plant sources, P for trend < 0.0001. Zheng et al. reported an odds ratio of 2.28, 95% confidence interval (1.71 , 3.04) for NDEA overall, P for trend < 0.0001.

Lung Cancer

14. Goodman et al.¹⁶ reported an odds ratio of 3.3, 95% confidence interval (1.7 , 6.2) for men and 2.7, 95% confidence interval (1.0 , 6.9) for women.
15. De Stefani et al.¹⁷ reported a lung cancer odds ratio of 3.14 (1.86 , 5.29).
16. Loh et al.¹⁸ reported a lung cancer hazard ratio of 1.05 (0.88 , 1.24)

Colorectal Cancer

¹² Cui, J., Guo, X. M., Bao, H. L., & Tan, J. B. (2016). Relationship between N-nitrosodimethylamine and risk of digestive tract cancers: a meta analysis based on cohort studies. *Chinese Journal of Endemiology*, 37(5), 725-729.

¹³ Rogers, M. A., Vaughan, T. L., Davis, S., & Thomas, D. B. (1995). Consumption of nitrate, nitrite, and nitrosodimethylamine and the risk of upper aerodigestive tract cancer. *Cancer Epidemiology and Prevention Biomarkers*, 4(1), 29-36.

¹⁴ R code is in Appendix A

¹⁵ Zheng, J., Stuff, J., Tang, H., Hassan, M. M., Daniel, C. R., & Li, D. (2019). Dietary N-nitroso compounds and risk of pancreatic cancer: results from a large case-control study. *Carcinogenesis*, 40(2), 254-262.

¹⁶ Goodman MT, Hankin JH, Wilkens LR, et al. High-fat foods and the risk of lung cancer. *Epidemiology* 1992;3:288-99.

¹⁷ De Stefani, E., Deneo-Pellegrini, H., Carzoglio, J. C., Ronco, A., & Mendilaharsu, M. (1996). Dietary nitrosodimethylamine and the risk of lung cancer: a case-control study from Uruguay. *Cancer Epidemiology and Prevention Biomarkers*, 5(9), 679-682.

¹⁸ Loh, Y. H., Jakszyn, P., Luben, R. N., Mulligan, A. A., Mitrou, P. N., & Khaw, K. T. (2011). N-nitroso compounds and cancer incidence: the European Prospective Investigation into Cancer and Nutrition (EPIC)-Norfolk Study. *The American Journal of Clinical Nutrition*, 93(5), 1053-1061.

17. Zhu et al.¹⁹ reported an odds ratio of 1.42, 95% confidence interval (1.03 , 1.96).
18. Knekt et al.²⁰ reported a colorectal cancer relative risk of 2.12 (1.04 , 4.33). Of note, the risk was increased for both the quartile 2 versus quartile 1 contrast and the quartile 3 versus quartile 1 contrast, albeit not statistically significant in either case. Since the mean NDMA level in Knekt et al. was 0.052 µg, this suggests increased risks at LCE's of less than 1,235 µg.
19. Loh et al.²¹ reported a colon cancer hazard ratio of 0.99 (0.83 , 1.18) and a rectum cancer hazard ratio of 1.46 (1.16 , 1.84).

Prostate Cancer

20. Loh et al.²² reported a prostate cancer hazard ratio of 1.01 (0.90 , 1.13).
21. Jakszyn et al.²³ reported a prostate cancer hazard ratio of 1.04 (0.92 , 1.18).

Bladder Cancer

22. Jakszyn et al.²⁴ reported a bladder cancer hazard ratio of 1.12 (0.88 , 1.44).
23. Table 1 below summarizes the individual studies. Despite the limitations of dietary studies such as measurement error that might introduce bias towards the null,²⁵ statistically significant associations and trends exist for all but bladder and prostate cancers. For gastric cancer, both the published meta-analysis and my meta-analysis show a statistically significant increased risk with the largest risks occurring in the two studies with the greatest lifetime

¹⁹ Zhu, Y., Wang, P. P., Zhao, J., Green, R., Sun, Z., Roebathan, B., ... & McLaughlin, J. R. (2014). Dietary N-nitroso compounds and risk of colorectal cancer: a case-control study in Newfoundland and Labrador and Ontario, Canada. *British Journal of Nutrition*, 111(6), 1109-1117.

²⁰ Knekt, P., Järvinen, R., Dich, J., & Hakulinen, T. (1999). Risk of colorectal and other gastro-intestinal cancers after exposure to nitrate, nitrite and N-nitroso compounds: a follow-up study. *International Journal of Cancer*, 80(6), 852-856.

²¹ Loh, Y. H., Jakszyn, P., Luben, R. N., Mulligan, A. A., Mitrou, P. N., & Khaw, K. T. (2011). N-nitroso compounds and cancer incidence: the European Prospective Investigation into Cancer and Nutrition (EPIC)-Norfolk Study. *The American Journal of Clinical Nutrition*, 93(5), 1053-1061.

²² Loh, Y. H., Jakszyn, P., Luben, R. N., Mulligan, A. A., Mitrou, P. N., & Khaw, K. T. (2011). N-nitroso compounds and cancer incidence: the European Prospective Investigation into Cancer and Nutrition (EPIC)-Norfolk Study. *The American Journal of Clinical Nutrition*, 93(5), 1053-1061.

²³ Jakszyn, P. G., Allen, N. E., Lujan-Barroso, L., Gonzalez, C. A., Key, T. J., Fonseca-Nunes, A., ... & Riboli, E. (2012). Nitrosamines and heme iron and risk of prostate cancer in the European prospective investigation into cancer and nutrition. *Cancer Epidemiology and Prevention Biomarkers*, 21(3), 547-551.

²⁴ Jakszyn, P., González, C. A., Luján-Barroso, L., Ros, M. M., Bueno-de-Mesquita, H. B., Roswall, N., ... & Riboli, E. (2011). Red meat, dietary nitrosamines, and heme iron and risk of bladder cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Cancer Epidemiology and Prevention Biomarkers*, 20(3), 555-559.

²⁵ Kipnis, V., Subar, A. F., Midthune, D., Freedman, L. S., Ballard-Barbash, R., Troiano, R. P., ... & Carroll, R. J. (2003). Structure of dietary measurement error: results of the OPEN biomarker study. *American Journal of Epidemiology*, 158(1), 14-21.

cumulative exposure. For gastric cancer, statistical significance arises at LCEs as low as 1,962 μg . I note that the daily NDMA exposure levels in the studies included in Song et al. are considerably lower than those arising with NDMA-contaminated valsartan (see Table 1 below).

24. For lung cancer, statistically significant increased risk arises with an LCE of 4,303 μg , and for rectal cancer with an LCE of 3,343 μg .
25. The Zhu et al. study has the second highest top-quintile level of NDMA LCE amongst these studies. Zhu et al. shows a statistically significant increased colorectal cancer risk when comparing the highest quintile to the lowest quintile as well as increased risk for other quintiles (LCEs of 1,002 μg , 3,030 μg , and 9,166 μg respectively), albeit not statistically significant.
26. Concerning NDEA, Zheng et al. shows a statistically significant increased risk at an LCE of 2,520 μg .

Table 1. Summary statistics from the studies.

First Author	Country	Design	Base High Dose μg	Approx. Average Age	LCE μg	Effect size	SS?	P for Trend
<i>Gastric</i>								
Palli	Italy	CC	0.24	60	5,260	1.1	N	0.7
DeStefani	Uruguay	CC	0.27	64	6,312	3.6	Y*	<0.001
Pobel	France	CC	0.27 [¶]	67	6,607	7.0	Y	0.04
LaVecchia	Italy	CC	0.19	51	3,539	1.4	Y	<0.01
Knekt	Finland	C	0.05 (mean)	41 + <24	-	0.8	N	0.4
Larsson	Sweden	C	0.19	~53 + 15	4,719	2.0	Y**	0.02
Keszei	Holland	C	0.15 ^M / 0.05 ^F ¶	61 + 16.3	4,235 ^M / 1,412 ^F	-	N	
Jakszyn06	EU	C	0.28 ^M / 0.11 ^F	51 + 6.6	5,891 ^M / 2,314 ^F	1.0	N	1.0
Loh	UK	C	0.13	59 + 11.4	3,343	1.1	N	0.5
<i>Esophagus</i>								
Keszei Sq.	Holland	C	0.15 ^M / 0.05 ^F ¶	61 + 16.3	4,235 ^M / 1,412 ^F	1.8	Y	0.01
Keszei Ad.	Holland	C	0.15 ^M / 0.05 ^F ¶	61 + 16.3	4,235 ^M / 1,412 ^F	-	N	-
Loh	UK	C	0.13	59 + 11.4	3,343	1.1	N	0.5
Rogers	US	CC	0.18	59	3,879	1.9	N	0.06
<i>Bladder</i>								
Jakszyn11	EU	C	0.19	51 + 8.7	4,143	1.1	N	0.5
<i>Prostate</i>								
Jakszyn12	EU	C	0.87	51 + 11	19,702	1.0	N	1.0
Loh	UK	C	0.13	59 + 11.4	3,343	1.0	N	0.9
<i>Pancreas</i>								
Zheng	US	CC	1.24	69	31,251	1.0	N	0.8
Zheng ^{NDEA}	US	CC	0.18	69	4,536	2.3	Y***	<0.0001
<i>Lung</i>								
DeStefani	Uruguay	CC	0.27	62	6,114	3.1	Y****	<0.001
Goodman ^M	US	CC	0.70	64	16,363	3.3	Y	0.0006
Goodman ^F	US	CC	0.70	64	16,363	2.7	Y	0.04
Loh	UK	C	0.13	59 + 11.4	3,343	1.1	N	0.6
<i>Colon/Colorectal</i>								
Knekt	Finland	C	0.05 (mean)	41 + <24	-	2.1	Y	0.5
Loh-Colon	UK	C	0.13	59 + 11.4	3,343	1.0	N	0.9
Loh-Rectal	UK	C	0.13	59 + 11.4	3,343	1.5	Y	0.001
Zhu	Canada	CC	1.24 [¶]	61	27,628	1.4	Y	0.005

* also significant for the second quantile – first quantile, LCE = 3,506 μg , and third quantile – first quantile contrasts, LCE = 4,441 μg .

** also significant for the third quantile – first quantile contrast, LCE = 1,962 μg

*** also significant for the second quantile – first quantile, LCE = 2,520 μg , and third quantile – first quantile contrasts, LCE = 3,024 μg .

**** also significant for the third quantile – first quantile contrast, LCE = 4,303 μg

¶ estimated – see Appendix A

4. Evidence from Occupational Studies

27. Hidajat et al. (2019) considers cancer mortality study in a cohort of 36,441 males aged 35+ years employed in the British rubber industry in 1967. The study followed the workers to 2015 with an average of 24.2 years per subject. Lifetime cumulative exposure (LCE) to NDMA was based on number of years worked and department. The exposure measurements derived from the EU-EXARUB database of measurements of compounds in rubber factories in Europe.
28. The primary analysis considered four different quartiles of NDMA LCE exposure defined as less than 3.12 year $\mu\text{g}/\text{m}^3$ (I), 3.12 - 5.96 year $\mu\text{g}/\text{m}^3$ (II), 5.96 – 9.67 year $\mu\text{g}/\text{m}^3$ (III), and greater than 9.67 year $\mu\text{g}/\text{m}^3$ (IV). Risk of death for all cancers comparing the highest NDMA exposure group (IV) to the lowest (I), yielded a hazard ratio of 2.08, and a corresponding 95% confidence interval of (1.96 , 2.21). Risk of death for all cancers comparing the next highest NDMA exposure group (III) to the lowest (I), yielded a hazard ratio of 1.83, 95% confidence interval (1.72 , 1.95). Finally, risk of death for all cancers comparing the next highest NDMA exposure group (II) to the lowest (I), yielded a hazard ratio of 1.32, 95% confidence interval (1.25 , 1.40). Assuming an average “active worker” breathes $10 \text{ m}^3/\text{day}$ ²⁶ and assuming 240 work days per year, this corresponds to 2.63 μg of NDMA per day. I derive this as follows:
- $9.67 / 24.2 = 0.40 \mu\text{g}/\text{m}^3$ is the average NDMA breathing rate.
 - $240 \times 10 = 2400$ is number of cubic meters breathed in a work year.
 - Thus $2400 \times 0.40 = 960$ is the number of μg of NDMA consumed per year.
 - $960/365.25 = 2.63 \mu\text{g}$ is the daily average per day.

Here is the calculation for each quartile:

Quartile	LCE	Daily equivalent	Total exposure
Q1	3.12 year $\mu\text{g}/\text{m}^3$	0.85 $\mu\text{g}/\text{day}$	7,513 μg
Q2	5.96 year $\mu\text{g}/\text{m}^3$	1.62 $\mu\text{g}/\text{day}$	14,319 μg
Q3	9.67 year $\mu\text{g}/\text{m}^3$	2.63 $\mu\text{g}/\text{day}$	23,247 μg

29. At a contamination level of 20 $\mu\text{g}/\text{day}$, the number of years of valsartan consumption needed to reach the total exposure equivalents in Hidajat is 1.0 year for Q1, 2.0 years for Q2, and 3.2 years for Q3.

²⁶ Galer, D. M., Leung, H. W., Sussman, R. G., & Trzos, R. J. (1992). Scientific and practical considerations for the development of occupational exposure limits (OELs) for chemical substances. *Regulatory Toxicology and Pharmacology*, 15(3), 291-306.

30. The Table below summarize Hidajat's findings with respect to individual cancers. Cells highlighted in yellow are statistically significant at the 5% level. The "P for Trend" is the p-value corresponding to a statistical test for increased risk associated with increased exposure. These results are conservative in the sense that they provide comparison with the lowest quartile of exposure, but the upper limit of that quartile is at LCE levels associated with increased risk in several dietary studies. They are also conservative in the sense that it is possible for those that are statistically significant at Q1 (respectively, Q2 and Q3), statistical significance could occur at levels lower than Q1 (respectively Q2 and Q3).

Cancer	Quartile			P for Trend
	Q1-Q2	Q2-Q3	>Q3	
Bladder	1.57 (1.19,2.07)	2.45 (1.87,3.21)	2.82 (2.16,3.67)	<0.01
Lung	1.21 (1.10,1.32)	1.54 (1.39,1.70)	1.70 (1.54,1.87)	0.36
Stomach	1.32 (1.10,1.57)	1.62 (1.32,1.98)	1.72 (1.41,2.10)	0.01
Leukemia	1.52 (0.99,2.33)	3.27 (2.20,4.86)	3.47 (2.35,5.13)	<0.01
M. Myeloma	1.59 (1.22,2.08)	2.78 (2.15,3.60)	2.81 (2.17,3.64)	<0.01
N.H. Lymphoma	1.51 (0.93,2.43)	2.17 (1.35,3.47)	2.25 (1.41,3.59)	0.11
Esophageal	1.70 (1.24,2.33)	2.43 (1.78,3.31)	3.04 (2.26,4.09)	0.26
Prostate	2.32 (1.82,2.97)	4.87 (3.89,6.11)	5.36 (4.27,6.73)	<0.01
Larynx	1.71 (0.93,3.14)	1.49 (0.69,3.19)	1.39 (0.67,2.90)	0.24
Brain	1.30 (0.76,2.25)	1.26 (0.68,2.36)	2.50 (1.53,-)	<0.01
Pancreas	1.59 (1.18,2.15)	2.19 (1.60,3.00)	2.60 (1.94,3.49)	0.42
Liver	1.53 (0.93,2.50)	1.96 (1.16,3.29)	2.86 (1.78,4.59)	0.03

31. I note that Hidajat et al. (2019) considered each contaminant (rubber dust, rubber fumes, nitrosamine sum score, NDMA, and NMor) separately so that the specific *cause* of the increased cancer mortality cannot be pinpointed. Hidajat et al (2020) considered multi-pollutant models for non-cancer mortality outcomes; the NDMA effects were attenuated but generally remained statistically significant.
32. Exposure in the Hidajat et al. study is via inhalation rather than ingestion, but Dr. Panigrahy has told me that NDMA and NDEA are similarly carcinogenic via either route (except that my calculations are conservative because Dr. Panigrahy explained that some amount of inhaled NDMA is exhaled).

5. Conclusion

33. For NDMA, statistically significant increased gastric cancer risk arises at LCEs as low as 1,962 µg. The equivalent threshold for lung cancer is 4,303 µg, for esophageal cancer is 4,235 µg, and for rectal cancer 3,343 µg. For NDEA, a statistically significant increased risk for pancreas cancer arises at an LCE of 2,520 µg.

34. Per Hidajat et al., cumulative exposure to greater than 7,514 μg of NDMA²⁷ statistically significantly increases one's risk of developing the following cancers – bladder, lung, stomach, multiple myeloma, esophageal, prostate, and prostate. Cumulative exposure greater than 14,319 μg of NDMA adds leukemia, lymphoma, and liver to the list. Cumulative exposure greater than 20,330 μg of NDMA adds brain cancer to the list.
35. Based on valsartan dosing, the levels of NDMA reported in contaminated valsartan and the timeframe over which the contamination occurred, it is scientifically plausible that users of contaminated valsartan could develop cancer. For example, in one year, a daily user of valsartan contaminated with 20 μg of NDMA would reach cumulative exposures for which both Hidajat et al. and several dietary studies show statistically significantly elevated risks of several cancers. Including NDMA from other sources would shorten this time further.

²⁷ $365.25 \times 24.4 \times 0.85$

APPENDIX A: R code

```

# combined Cui and Rogers
library(metafor)
esoph <- data.frame(
  group = factor(c("Cui", "Rogers")),
  or = log(c(1.18, 1.86)),
  lcl = log(c(0.98, 0.87)),
  ucl = log(c(1.41, 3.95)))
esoph$vi <- ((esoph$ucl - esoph$lcl) / (2*1.96))^2
res <- rma(or, vi, data=esoph, method="DL")

# extend Song to include Loh
library(metafor)
gastro <- data.frame(
  group = factor(c("Knekt", "Jakszyn", "Larsson", "KeszeiWC", "KeszeiWN", "KeszeiMC",
    "KeszeiMN", "LaVecchia", "Pobel", "DeStefani", "Palli", "Loh")),
  or = log(c(0.75, 0.99, 1.96, 1.02, 0.90, 0.94, 1.31, 1.37, 7.00, 3.62, 1.10, 1.13)),
  lcl = log(c(0.37, 0.69, 1.08, 0.33, 0.58, 0.59, 0.95, 1.1, 1.85, 2.38, 0.8, 0.81)),
  ucl = log(c(1.51, 1.41, 3.58, 3.14, 1.42, 1.49, 1.81, 1.7, 26.46, 5.51, 1.5, 1.57)))
gastro$vi <- ((gastro$ucl - gastro$lcl) / (2*1.96))^2
res <- rma(or, vi, data=gastro, method="DL")

# Q3 calculation for Pobel
x<-c(25,50,75)
y<-c(0.19,0.23,0.29)
M<-lm(log(y)~x)
exp(coef(M)[1]+(66.67*coef(M)[2]))

# Q5 calculation for Zhu
y<-c(0.03,0.07,0.2,0.77,2.29)
x<-c(10,30,50,70,90)
M<-lm(log(y)~x)
exp(coef(M)[1]+(80*coef(M)[2]))

# Q3 calculation for Keszei
x<-c(16.67,50,83.3)
#y<-c(0.04,0.08,0.25)
y<-c(0.03,0.04,0.07)
M<-lm(log(y)~x)
exp(coef(M)[1]+(66.67*coef(M)[2]))

```

APPENDIX 1

David Madigan

davidbennettmadigan@gmail.com

Tel: (862) 812-3690

Curriculum Vitae**1 June 2021****Education**

Trinity College Dublin, Ph.D., Statistics, 1990. Dissertation "An investigation of weights of evidence in the context of probabilistic expert systems." K. R. Mosurski, Advisor.

Trinity College Dublin, B.A. (Mod.), Mathematics, 1984, First Class Honours.

Employment History

2020 - : Northeastern University

2020 - : Provost & Senior Vice-President for Academic Affairs

2020 - : Professor of Statistics, Khoury College of Computer Sciences

2007 - 2020 : Columbia University

2007 - 2020 : Professor of Statistics

2013 - 2018 : Executive Vice-President for Arts and Sciences

2013 - 2018 : Dean of the Faculty of Arts and Sciences

2007 - 2013 : Chair, Department of Statistics

2001 - 2007 : Rutgers University

2001 - 2007 : Professor of Statistics and Biostatistics

2005 - 2007 : Dean, Physical and Mathematical Sciences

2003 - 2004 : Director, Institute of Biostatistics

2000 - 2001 : Vice President, Data Mining, Soliloquy, Inc.

1999 - 2000 : Principal Technical Staff Member, AT&T Labs-Research

1990 - 1999 : University of Washington/ Fred Hutchinson Cancer Research Center

1995 - 1999 : Associate Professor of Statistics, UW

1992 - 1999 : Assistant/Associate Member, FHCRC

1990 - 1995 : Assistant Professor of Statistics, UW

1989 - 1990 : Information Technology Consultant, KPMG, Ireland

1986 - 1989 : Technology Manager, Peregrine Expert Systems Ltd., Ireland

1985 - 1986 : Expert System Consultant, SkillSoft, Ireland

1984 - 1985 : Actuarial Associate, Hibernian Life Assurance, Ireland

Honors

2014: Elected Member of the International Statistical Institute
 2012: Elected Fellow of the American Association for the Advancement of Science.
 2009: Institute of Mathematical Statistics Medallion Lecturer.
 2006: Elected Fellow of the Institute of Mathematical Statistics.
 2005: 36th Most Cited Mathematician in the World, 1995-2005, ISI Thomson.
 1999: Elected Fellow of the American Statistical Association.
 1995: University of Washington Distinguished Teaching Award.
 1984: Gold medal awarded by the board of Trinity College Dublin.
 1980: Trinity College Dublin, Entrance Scholarship in Mathematics.

Refereed Publications

1. Chen, R., Suchard, M.A., Krumholz, H.M., Schuemie, M.J., Shea, S., Duke, J., Pratt, N., Reich, C.G., Madigan, D., You, S.C., Ryan, P.B., and Hripcsak, G., (2021). Comparative first-line effectiveness and safety of angiotensin converting enzyme inhibitors and angiotensin receptor blockers: a multinational cohort study. *Hypertension*, to appear.
2. Hripcsak, G., Schuemie, M.J., Madigan, D., Ryan, P.B., and Suchard, M. (2021). Drawing reproducible conclusions from observational clinical data with OHDSI. *Yearbook of Medical Informatics*, DOI: 10.1055/s-0041-1726481.
3. Park, S., You, S.C., Krumholz, H.M., Suchard, M.A., Schuemie, M., Hripcsak, G., Chen, R., Shea, S., Duke, J., Pratt, N., Reich, C., Madigan, D., Ryan, P., and Park, R.W. (2021). Comprehensive comparative effectiveness and safety of first-line beta-blocker monotherapy in hypertensive patients: a large-scale multi-center observational study. *Hypertension*, to appear.
4. Dwivedi, R., Tan, Y. S., Park, B., Wei, M., Horgan, K., Madigan, D., & Yu, B. (2020). Stable discovery of interpretable subgroups via calibration in causal studies. *International Statistical Review*, 88, S1, S135-S178 doi:10.1111/insr.12427.
5. You, S.C., Rho, Y., Bikdeli, B., Kim, J., Siapos, A., Weaver, J., Londhe, A., Cho, J., Park, J., Schuemie, M., Suchard, M.A., Madigan, D., Hripcsak, G., Gupta, A., Reich, C. G., Ryan, P.B., Park, R.W., and Krumholz, H.M. (2020). Association of ticagrelor versus clopidogrel with net adverse clinical events in patients with acute coronary syndrome undergoing percutaneous coronary intervention. *JAMA*. 2020;324(16):1640-1650. doi:10.1001/jama.2020.16167.
6. Kim, Y., Tian, Y., Yang, J., Huser, V., Jin, P., Lambert, C., Park, H., You, S.C., Park, R.W., Rijnbeek, P., Zandt, M., Reich, C., Vashisht, R., Wu, Y., Duke, J., Hripcsak, G., Madigan, D., Shah, N., Ryan, P., Schuemie, M., Suchard, M. (2020). Comparative safety and effectiveness of alendronate versus raloxifene in women with osteoporosis. *Scientific Reports*, 10.1 (2020): 1-10.
7. Schuemie, M.J., Ryan, P.B., Pratt, N., You, S.C., Krumholz, H.M., Madigan, D., Hripcsak, G. and Suchard, M.A. (2020). Principles of Large-Scale Evidence Generation and Evaluation

- across a Network of Databases (LEGEND). *Journal of the American Medical Informatics Association*, <https://doi.org/10.1093/jamia/ocaa103>
8. Schuemie, M.J., Ryan, P.B., Pratt, N., You, S.C., Krumholz, H.M., Madigan, D., Hripcsak, G. and Suchard, M.A. (2020). Large-Scale Evidence Generation and Evaluation across a Network of Databases (LEGEND): Assessing Validity Using Hypertension as a Case Study. *Journal of the American Medical Informatics Association*, <https://doi.org/10.1093/jamia/ocaa124>.
 9. Schuemie, M.J., Cepeda, M.S., Suchard, M.A., Yang, J., Tian, Y., Schuler, A., Ryan, P.B., Madigan, D., and Hripcsak, G. (2020). How Confident Are We About Observational Findings in Healthcare: A Benchmark Study. *Harvard Data Science Review*, 2.1, DOI: 10.1162/99608f92.147cc28e.
 10. Hripcsak, G., Suchard, M.A., Shea, S., Chen, R., Pratt, N., Madigan, D., Krumholz, H.M., Ryan, P.B., and Schuemie, M.J. (2019). Real-World Evidence on the Effectiveness and Safety of Chlorthalidone and Hydrochlorothiazide. *JAMA Internal Medicine*, doi:10.1001/jamainternmed.2019.7454.
 11. Suchard, M.A., Schuemie, M.J., Krumholz, H.M., You, S., Chen, R., Pratt, N., Reich, C.G., Duke, J., Madigan, D., Hripcsak, G., and Ryan, P.B. (2019). Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes. *The Lancet*, DOI:[https://doi.org/10.1016/S0140-6736\(19\)32317-7](https://doi.org/10.1016/S0140-6736(19)32317-7). *International Medical Informatics Association Best Paper 2019 - Bioinformatics and Translational Informatics*.
 12. Lu, F., Zheng, Y., Cleveland, H., Burton, C., and Madigan, D. (2018). Bayesian hierarchical vector autoregressive models for patient-level predictive modeling. *PLoS ONE*, <https://doi.org/10.1371/journal.pone.0208082>.
 13. Schuemie, M., Ryan, P., Hripcsak, G., Madigan, D., and Suchard, M. (2018). Improving reproducibility by using high-throughput observational studies with empirical calibration. *Philosophical Transactions A*, **376**:20170356. <http://dx.doi.org/10.1098/rsta.2017.0356>.
 14. Madigan, D. and Shin, J. (2018). Drospirenone-Containing Oral Contraceptives and Venous Thromboembolism: An Analysis of the FAERS Database. *Open Access Journal of Contraception*, **9**:29-32.
 15. Schuemie, M.J., Ryan, P., Hripcsak, G., Madigan, D., and Suchard, M. (2018). Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proceedings of the National Academy of Science*, <https://doi.org/10.1073/pnas.1708282114>.
 16. Berger, M., Sox, H., Willke, R., Brixner, D., Eichler, H-G., Madigan, D., Makady, A., Schneeweiss, S., Tarricone, R., Wang, S., Mullins, D., Watkins, J. (2017). Recommendations for Good Procedural Practices for Real-World Data Studies of Treatment Effectiveness and/or Comparative Effectiveness: Report of the Joint ISPOR-ISPE Special Task Force on Real-World Evidence in Health Care Decision Making. *Pharmacoepidemiology and Drug Safety*, DOI: 10.1002/pds.4297.
 17. Zannad, F., Michael S. Lauer, Robert Temple, Marc A. Pfeffer, Deepak L. Bhatt, Denise E. Bonds, Jeffrey S. Borer, Gonzalo Calvo, Louis Fiore, Lars H. Lund, David Madigan, Aldo P. Maggioni, Jerry A. Menikoff, Catherine M. Meyers, Yves Rosenberg, Tabassome Simon, Wendy Gattis Stough, Andrew Zalewski, Nevine Zariffa, Robert M. Califf (2017). Streamlining Cardiovascular Clinical Trials to Achieve Improved Efficiency and Generalizability: Current Progress and Future Steps. *Heart*, to appear.

18. Sobel, M., Madigan, D., and Wang, W. (2017). Meta-analysis: A causal framework, with application to randomized studies of Vioxx. *Psychometrika*. 82: 459, 10.1007/s11336-016-9507-z
19. Shahn, Z. and Madigan, D. (2016). Latent Class Mixture Models of Treatment Effect Heterogeneity. *Bayesian Analysis*, DOI: 10.1214/16-BA1022.
20. Shaddox, T.R., Ryan, P.B., Schuemie, M.J., Madigan, D., and Suchard, M.A. (2016). Hierarchical models for multiple rare outcomes using massive observational healthcare databases. *Statistical Analysis and Data Mining*, 9(4), 260-268.
21. Selzman, C. H., Felker, E., Sheridan, B. C., Silvestry, S., Daly, R. C., Anyanwu, A., Madigan, D., Frazier, O. & Griffith, B. P. (2016). The Jarvik 2000: Results of the United States Bridge to Transplant Trial. *The Journal of Heart and Lung Transplantation*, 35(4), S38.
22. Webman, R. B., Carter, E. A., Mittal, S., Wang, J., Sathya, C., Nathens, A. B., Nance, M.L., Madigan, D. & Burd, R. S. (2016). Association between trauma center type and mortality among injured adolescent patients. *JAMA Pediatrics*, 170(8), 780-786.
23. Hripcsak, G., Patrick Ryan, Jon Duke, Nigam H. Shah, Rae Woong Park, Vojtech Huserh, Marc A. Suchard, Martijn Schuemie, Frank DeFalco, Adler Perotte, Juan Banda, Christian Reich, Lisa Schilling, Michael Matheny, Daniella Meeker, Nicole Pratt, and Madigan, D. (2016). Addressing Clinical Questions at Scale: OHDSI Characterization of Treatment Pathways. *Proceedings of the National Academy of Sciences*, vol. 113 no. 277329–7336, doi: 10.1073/pnas.1510502113.
24. Moghaddass, R., Rudin, C., and Madigan, D. (2016). The Factorized Self-Controlled Case Series Method: An Approach for Estimating the Effects of Many Drugs on Many Outcomes. *Journal of Machine Learning Research*, 17(185):1–24, 2016.
25. Beck, H.E., Mittal, S., Madigan, D., and Burd, R.S. (2015). Reassessing mechanism as a predictor of pediatric injury mortality. *Journal of Surgical Research* 199 (2), 641-646.
26. Boland, M.R., Z Shahn, D Madigan, G Hripcsak, NP Tatonetti (2015). Birth Month Affects Lifetime Disease Risk: A Phenome-Wide Method. *Journal of the American Medical Informatics Association*, DOI: <http://dx.doi.org/10.1093/jamia/ocv046>.
27. Shahn, Z., Ryan, P., and Madigan, D. (2015). Predicting Health Outcomes from High Dimensional Longitudinal Health Histories Using Relational Random Forests. *Statistical Analysis and Data Mining*, 8:128-136, DOI: 10.1002/sam.11268.
28. Berger, M.L., Lipset, C., Gutteridge, A., Axelsen, K., Subedi, P., and Madigan, D. (2015). Optimizing the Leveraging of Real World Data: How It Can Improve the Development and Use of Medicines? *Value in Health*, <http://dx.doi.org/10.1016/j.jval.2014.10.009>.
29. Hripcsak, G., Jon D Duke, Nigam H Shah, Christian G Reich, Vojtech Huser, Martijn J Schuemie, Marc A Suchard, Rae Woong Park, Ian Chi Kei Wong, Peter R Rijnbeek, Johan van der Lei, Nicole Pratt, G Niklas Norén, Yu-Chuan Li, Paul E Stang, David Madigan, and Patrick B Ryan (2015). Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *MedInfo, Stud Health Technol Inform*. 2015; 216: 574–578.
30. Letham, B., Rudin, C., McCormick, T.H., and Madigan, D. (2014). Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model. *Annals of Applied Statistics*, 9, 1350-1371, DOI: 10.1214/15-AOAS848.
31. Hripcsak, G., Varela, S.V., Ryan, P.B., Madigan, D., Stang, P., Schuemie, M., Friedman, C., and Tatonetti, N. (2014). Similarity-based Modeling applied to Signal Detection in Pharmacovigilance. *CPT: Pharmacometrics & Systems Pharmacology*, 3(9), 1-9.

32. Schuemie, M.J., Trifirò, G., Coloma, P.M., Ryan, P.B. and Madigan, D. (2014). Detecting adverse drug reactions following long-term exposure in longitudinal observational data. *Statistical Methods in Medical Research*, doi:10.1177/0962280214527531.
33. Price, K. L., Xia, H.A., Lakshminarayanan, M., Madigan, D., Manner, D., Scott, J., Stamey, J., Thompson, L. (2014). Bayesian Methods for Design and Analysis of Safety Trials. *Pharmaceutical Statistics*, 13, 13-24.
34. Simpson, S., Madigan, D., Zorych, I., Schuemie, M.J., Ryan, P.B., and Suchard, M. (2013). Multiple self-controlled case series for large-scale longitudinal observational databases. *Biometrics*, DOI: 10.1111/biom.12078.
35. Ryan, P.B., Schuemie, M.J., Gruber, S., Zorych, I., and Madigan, D. (2013). Empirical Performance of a New User Cohort Method: Lessons for Developing a Risk Identification and Analysis System. *Drug Safety*, **36** (Suppl 1):S59-S72.
36. Madigan, D., Schuemie, M.J., and Ryan, P.B. (2013). Empirical Performance of the Case-Control Method: Lessons for Developing a Risk Identification and Analysis System. *Drug Safety*, **36** (Suppl 1):S73-S82.
37. Suchard, M.A., Zorych, I., Simpson, S.E., Schuemie, M.J., Ryan, P.B., and Madigan, D. (2013). Empirical Performance of the Self-Controlled Case Series Design: Lessons for Developing a Risk Identification and Analysis System. *Drug Safety*, **36** (Suppl 1):S83-S93.
38. Ryan, P.B., Schuemie, M.J., and Madigan, D. (2013). Empirical Performance of a Self-Controlled Cohort Method: Lessons for Developing a Risk Identification and Analysis System. *Drug Safety*, **36** (Suppl 1):S95-S106.
39. Schuemie, M.J., Madigan, D., and Ryan, P.B. (2013). Empirical Performance of LGPS and LEOPARD: Lessons for Developing a Risk Identification and Analysis System. *Drug Safety*, **36** (Suppl 1):S133-S142.
40. Noren, G.N., Bergvall, T., Ryan, P.B., Juhlin, K., Schuemie, M.J., and Madigan, D. (2013). Empirical performance of the calibrated self-controlled cohort analysis within Temporal Pattern Discovery: Lessons for developing a risk identification and analysis system. *Drug Safety*, **36** (Suppl 1):S107-S121.
41. Ryan, P.B., Stang, P.E., Overhage, J.M., Suchard, M.A., Hartzema, A.G., DuMouchel, W., Reich, C.G., Schuemie, M.J., and Madigan, D. (2013). A Comparison of the Empirical Performance of Methods for a Risk Identification System. *Drug Safety*, **36** (Suppl 1):S143-S158.
42. DuMouchel, W., Ryan, P.B., Schuemie, M.J., and Madigan, D. (2013). Evaluation of disproportionality safety signalling applied to healthcare databases. *Drug Safety*, **36** (Suppl 1):S123-S132.
43. Hartzema, A.G., Reich, C.G., Ryan, P.B., Stang, P.E., Madigan, D., Welebob, E., Overhage, J.M. (2013). Managing data quality for a drug safety surveillance system. *Drug Safety*, **36** (Suppl 1):S49-S58.
44. Ryan, P.B., Madigan, D., Stang, P.E., Schuemie, M.J., and Hripcsak, G. (2013). Medication-wide association studies. *CPT: Pharmacometrics & Systems Pharmacology* **2**, e76; doi:10.1038/psp.2013.52.
45. Mittal, S., Madigan, D., Suchard, M., and Burd, R. (2013). High-Dimensional, Massive Sample-Size Cox Proportional Hazards Regression for Survival Analysis. *Biostatistics*, doi: 10.1093/biostatistics/kxt043.

46. Rudin, C., Letham, B., and Madigan, D. (2013). Learning theory analysis for association rules and sequential event prediction. *Journal of Machine Learning Research*, **14**, 3441-3492.
47. Madigan, D., Stang, P.E., Berlin, J.A., Schuemie, M.J., Overhage, J.M., Suchard, M.A., DuMouchel, W., Hartzema, A.G., and Ryan P.B. (2013). A Systematic Statistical Approach to Integrating Information from Observational Studies. *Annual Review of Statistics and Its Application*, **1**, 11-39.
48. Schuemie, M., Ryan, P., DuMouchel, W., Suchard, M.A., and Madigan, D. (2013). Interpreting observational studies - why empirical calibration is needed to correct p-values. *Statistics in Medicine*, DOI: 10.1002/sim.5925.
49. Simpson, S., Madigan, D., Zorych, I., Schuemie, M.J., Ryan, P.B., and Suchard, M. (2013). Multiple self-controlled case series for large-scale longitudinal observational databases. *Biometrics*, **69**, 893-902.
50. Letham, B., Rudin, C., McCormick, T.H., and Madigan, D. (2013). An interpretable stroke prediction model with using rules using Bayesian analysis. *Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI-13) Late Breaking Paper*.
51. Letham, B., Rudin, C., and Madigan, D. (2013). A supervised ranking approach to sequential event prediction. *Machine Learning*, 10.1007/s10994-013-5356-5.
52. Ryan, P., Suchard, M.A., Schuemie, M., and Madigan, D. (2013). Learning from epidemiology: Interpreting observational studies for the effects of medical products. *Statistics in Biopharmaceutical Research*, DOI:10.1080/19466315.2013.791638.
53. Mittal, S., Madigan, D., Cheng, J., and Burd, R. (2013). Large-scale Bayesian parametric survival analysis. *Statistics in Medicine*, DOI: 10.1002/sim.5817.
54. Emir, B., Amaratunga, D., Beltangady, M., Cabrera, J., Freeman, R., Madigan, D., Nguyen, H., and Whalen, E. (2013). Generating productive dialogue between consulting statisticians and their clients ion the pharmaceutical and medical research settings. *Open Access Medical Statistics*, **3**, 51-56.
55. Madigan, D., Ryan, P., Schuemie, M., Stang, P., Overhage, M., Hartzema, A., Suchard, M.A., DuMouchel, W., and Berlin, J. (2013). Evaluating the impact of database heterogeneity on observational studies. *American Journal of Epidemiology*, DOI: 10.1093/aje/kwt010.
56. Suchard, M., Simpson, S.E., Zorych, I., Ryan, P., and Madigan, D. (2013). Massive parallelization of serial inference algorithms for generalized linear models. *ACM Transactions on Modeling and Computer Simulation*, **23**:1-17.
57. Madigan, D., Ryan, P.B., and Schuemie, M.J. (2012). Does design matter? Systematic evaluation of the impact of analytical choices on effect estimates in observational studies. *Therapeutic Advances in Drug Safety*, **4**, 53-62.
58. Ryan, P.B., Madigan, D., Stang, P.E., Overhage, J.M., Racoosin, J.A., Hartzema, A.G. (2012). Empirical Assessment of Analytic Methods for Risk Identification in Observational Healthcare Data: Results from the Experiments of the Observational Medical Outcomes Partnership. *Statistics in Medicine*, **30**, 4401-4415.
59. Madigan, D., Sigelman, D., Mayer, J.W., Furberg, C.D., Avorn, J. (2012). Under-reporting of cardiovascular events in the rofecoxib Alzheimer studies. *American Heart Journal*, doi:10.1016/j.ahj.2012.05.002.

60. Harpaz, R., DuMouchel, W., Shah, N.H., Madigan, D., Ryan, P., and Friedman, C. (2012). Novel data mining methodologies for adverse drug event discovery and analysis. *Clinical Pharmacology & Therapeutics*, doi:10.1038/clpt.2012.50.
61. Maclure, M., Fireman, B., Nelson, J.C., Hua, W., Shoaibi, A., Paredes, A., and Madigan, D. (2012). When should a distributed system for active medical product surveillance use case-based designs for safety monitoring?. *Pharmacoepidemiology and Drug Safety*, **21**, 50-61.
62. McCormick, T., Madigan, D., Raftery, A.E., and Burd, R.S. (2012). Dynamic model averaging for logistic regression. *Biometrics*, **68**:23-30.
63. McCormick, T., Rudin, C., and Madigan, D. (2012). A hierarchical model for association rule mining of sequential events: an approach to automated medical symptom prediction. *Annals of Applied Statistics*, **6**, 652-658.
64. Oquendo, M.A., Baca-Garcia, E., Artes, A., Perez-Cruz, F., Galfalvy H.C., Blasco-Fontecilla, H., Madigan, D., Duan, N. (2011). Hypothesis Generation in the 21st Century. *Molecular Psychiatry*, to appear.
65. Smith, R.T., Merriam, J.E., Sohrab, M.A., Pumariega, N.M., Barile, G., Blonska, A.M., Haans, R., Madigan, D., and Allikmets, R. (2011). Complement Factor H 402H Variant and Reticular Macular Disease. *Archives of Ophthalmology* 129(8):1061-1066.
66. Madigan, D., Mittal, S., and Roberts, F. (2011). Efficient sequential decision making algorithms for container inspection operations. *Naval Research Logistics*, **58**, 637-654.
67. Zorych, I., Madigan, D., Ryan, P., and Bate, A. (2011). Disproportionality methods for pharmacovigilance in longitudinal observational databases. *Statistical Methods in Medical Research*, doi: 10.1177/0962280211403602.
68. Rudin, C., Salleb-Aouissi, A., Kogan, E. and Madigan, D. (2011). Sequential Event Prediction with Association Rules. *Proceedings of the 2011 Conference on Learning Theory (COLT)* (30%). Also JMLR: Workshop and Conference Proceedings 19 (2011) 615-634.
69. Madigan, D. and Ryan, P. (2011). What can we really learn from observational studies? The need for empirical assessment of methodology for active drug safety surveillance and comparative effectiveness research. *Epidemiology*, **22** (5), 629-631.
70. Madigan, D., Ryan, P., Simpson, S.E., and Zorych, I. (2010). Bayesian methods in pharmacovigilance (with discussion). In: J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West (eds), *Bayesian Statistics 9*, Oxford University Press, 421-438.
71. Balakrishnan, S. and Madigan, D. (2010). Priors on the variance in sparse Bayesian learning: the demi-Bayesian lasso. In: *Frontiers of Statistical Decision Making and Bayesian Analysis: In Honor of James O. Berger* by Ming-Hui Chen, Peter Müller, Dongchu Sun, and Keying Ye.
72. Caster, O., Noren, G.N., Madigan, D., and Bate, A. (2010). Large-Scale Regression-Based Pattern Discovery: The Example of Screening the WHO Global Drug Safety Database. *Statistical Analysis and Data Mining*, **3**, 197-208.
73. Ross, J.S., Madigan, D., Konstam, M.A., Egilman, D.S., and Krumholz, H.M. (2010). Does Rofecoxib cardiovascular risk persist after discontinuation? *Archives of Internal Medicine*, **170**, 2035-2036.

74. Cheng, J. and Madigan, D. (2010). Bayesian Approaches to Aspects of the Vioxx Trials: Non-ignorable Dropout and Sequential Meta-Analysis. In: *Handbook of Applied Bayesian Analysis*, Oxford University Press, 51-68.
75. Ross, J.S., Madigan, D., Hill, K.P., Egilman, D.S., Wang, Y., Krumholz, H.M. (2009). Pooled analysis of Rofecoxib placebo-controlled clinical trial data: Lessons for post-market pharmaceutical safety surveillance. *Archives of Internal Medicine*, **169**, 1976-1985.
76. Burd, R. and Madigan, D. (2009). An evaluation of the impact of injury coding schemes on mortality prediction in pediatric trauma. *Academic Emergency Medicine*, **16**, 639-645.
77. Pearson, R.K., Hauben, M., Goldsmith, D., Gould, A.L., Madigan, D., O'Hara, D.J., Reisinger, S., and Hochberg, A. (2009). Influence of the MEDDRA hierarchy on pharmacovigilance data mining results. *International Journal of Medical Informatics*, **78**, e97-e103.
78. Hochberg, A., Hauben, M., Pearson, R.K., O'Hara, D., Reisinger, S., Goldsmith, D.I., Gould, A.L., and Madigan, D. (2009). An Evaluation of Three Signal Detection Algorithms Using a Highly Inclusive Reference Event Database. *Drug Safety*, **32**, 509-525.
79. Burd, R.S., Ouyang, M., and Madigan, D. (2008). Bayesian logistic injury severity score (BLISS): A method for predicting mortality using ICD-9 codes. *Academic Emergency Medicine*, **15**(5), 466-475.
80. Caster, O., Noren, G.N., Madigan, D., and Bate, A. (2008). Large-scale regression-based pattern discovery: The example of the WHO drug safety database. *KDD Workshop on Mining Medical Data*, to appear.
81. Naik, P., Wedel, M., Bacon, L., Bodapati, A., Bradlow, E., Kamakura, W., Kreulen, J., Lenk, P., Madigan, D., and Montgomery, A. (2008). Challenges and Opportunities in High-Dimensional Choice Data Analysis." *Marketing Letters*, **19** (3), 201-213.
82. Balakrishnan, S. and Madigan, D. (2007). Algorithms for Sparse Linear Classifiers in the Massive Data Setting. *Journal of Machine Learning Research*, **9**, 313-337, 2007.
83. Balakrishnan, S. and Madigan, D. (2007). LAPS: Lasso with Partition Search. *Proceedings of the IEEE International Conference on Data Mining*, 415-420, 19% acceptance rate.
84. Hauben, M., Madigan, D., Reisinger, S., Hochberg, A., and O'Hara, D. (2007). Data Mining in Pharmacovigilance: Computational Cost as a Neglected Performance Parameter. *International Journal of Pharmaceutical Medicine*, **21**, 319-323.
85. Hauben, M., Reich, L., Gerrits, C.M., and Madigan, D. (2007). Spontaneous Reporting of Hyperkalemia and The Randomized Aldactone Evaluation Study. *Drug Safety*, **30**, 1143-1149.
86. Eyheramendy, D. and Madigan, D. (2007). A Bayesian feature selection score based on Naive Bayes models. In: *Computational Methods of Feature Selection*, H. Liu and H. Motoda, Editors, 277-294.
87. Madigan, D., Mittal, S., and Roberts, F. (2007). Sequential decision making algorithms for port of entry inspection: overcoming computational challenges. *Proceedings of 2007 Intelligence and Security Informatics Conference*, 1-7.
88. Eyheramendy, S. and Madigan, D. (2007). A Flexible Bayesian Generalized Linear Model for Dichotomous Response Data with an Application to Text Categorization. In: *IMS Lecture Notes - Monograph Series, Volume 54, Complex datasets and inverse problems: tomography, networks and beyond*. Regina Liu, William Strawderman & Cun-Hui Zhang, Editors, 76-91.

89. Genkin, A., Lewis, D.D., and Madigan, D. (2007). Large-scale Bayesian logistic regression for text categorization. *Technometrics*, **49**, 291-304.
90. Rolka, H., Burkom, H., Cooper, G.F., Kulldorff, M., Madigan, D., and Wong, W-K. (2006). Issues in Applied Statistics for Public Health Bioterrorism Surveillance using Multiple Data Streams: Research needs. *Statistics in Medicine*, **26**, 1834-1856.
91. Balakrishnan, S. and Madigan, D. (2006). Decision Trees for Functional Variables. *Proceedings of the IEEE International Conference on Data Mining*, 798-802, 20% acceptance rate.
92. Dayanik, A., Lewis, D.D., Madigan, D., Menkov, V., and Genkin, A. (2006). Constructing Informative Prior Distributions from Domain Knowledge in Text Classification, *Proceedings of the 29th Annual International ACM SIGIR conference*, 493-500 (18.5% acceptance rate).
93. Anand, S., Madigan, D., Mammone, R., Pathak, S. and Roberts, F. (2006). Experimental Analysis of Sequential Decision Making Algorithms for Port of Entry Inspection Procedures. In S. Mehrotra, D. Zeng, H. Chen, B. Thuraisingham, and F-X Wang (eds.), *Intelligence and Security Informatics, Proceedings of ISI-2006, Lecture Notes in Computer Science #3975*, Springer-Verlag, New York, 2006.
94. Balakrishnan, S. and Madigan, D. (2006). A One-Pass Sequential Monte Carlo Method for Bayesian Analysis of Massive Datasets. *Bayesian Analysis*. **1**, 345-362.
95. Madigan, D., Ju, W., Krishnan, P., and Krishnakumar, A.S. (2006). Location estimation in wireless networks: A Bayesian approach. *Statistica Sinica*, **16**, 495-522.
96. Madigan, D., Vardi, Y., and Weissman, I. (2006). Extreme value theory applied to document retrieval from large collections. *Information Retrieval*, **9**, 273-294.
97. Hauben, M., Madigan, D., Gerrits, C.M., Walsh, L., and Van Puijenbroek, E.P. (2005). The role of data mining in pharmacovigilance. *Expert Opinion in Drug Safety*., **4**(5), 929-948.
98. Madigan, D., Genkin, A., Lewis, D.D., and Fradkin, D. (2005). Bayesian multinomial logistic regression for author identification. *Proceedings of the 25th International Workshop on Bayesian inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 05)*, 509-516.
99. Madigan, D., Genkin, A., Argamon, S., Fradkin, D., and Ye, L. (2005). Author identification. *Proceedings of CSNA/Interface 05*.
100. Eyheramendy, S. and Madigan, D. (2005). A Novel Feature Selection Score for Text Categorization. *International Workshop on Feature Selection for Data Mining*, 1-8, 24% acceptance rate.
101. Madigan, D. (2005). Statistics and Data Mining. In: *AMS-DIMACS Discrete Methods in Epidemiology*, James Abello and Graham Cormode (Editors), 21-24.
102. Hanks, S. and Madigan, D. (2005). Probabilistic temporal reasoning. In: *Handbook of Temporal Reasoning in Artificial Intelligence*, M. Fisher, D. Gabbay, and L. Vila, Editors, Elsevier B.V., 315-342.
103. Madigan, D. (2005). Bayesian data mining for surveillance. In: *Spatial and Syndromic Surveillance for Public Health* (Andrew Lawson and Ken Kleinman, Editors), 203-221.
104. Madigan, D., Elnahrawy, E., Martin, R.P., Ju, W., Krishnan, P. and Krishnakumar, A.S. (2004). Bayesian Indoor Positioning Systems. *Proceedings of IEEE Infocom*, 1217-1227 (17% acceptance rate)

105. Madigan, D. (2004). Statistics and the war on spam. In: *Statistics: A Guide to the Unknown*, Deb Nolan (Editor), 135-147.
106. Fradkin, D. and Madigan, D.. (2003). Experiments with random projections for machine learning. In *Proceedings of KDD-03, The Ninth International Conference on Knowledge Discovery and Data Mining*, 517-522.
107. Ridgeway, G. and Madigan, D. (2003). A sequential Monte Carlo Method for Bayesian analysis of massive datasets. *Journal of Knowledge Discovery and Data Mining*, **7**, 301-319.
108. Eyheramendy, S., Lewis, David D., and Madigan, David (2003). On the naïve bayes model for text classification. In *Proceedings of The Ninth International Workshop on Artificial Intelligence and Statistics*, C.M. Bishop and B.J. Frey (Editors), 332-339.
109. Madigan, D., Vardi, Y., and Weissman, I. (2003). On retrieval properties of samples of large collections. In *Proceedings of The Ninth International Workshop on Artificial Intelligence and Statistics*, C.M. Bishop and B.J. Frey (Editors), 265-270.
110. Madigan, D. and Ridgeway, G. (2003). Bayesian data analysis for data mining. In *Handbook of Data Mining*, N. Ye (Ed.), 103-132..
111. Cohen, A., Madigan, D., and Sackrowitz, H.B. (2003). Effective directed tests for models with ordered categorical data. *Australian and New Zealand Journal of Statistics*, **45**, 285-200. 2003 Best Paper Award.
112. Mangione, S., Yuen, E., and Madigan, D. (2003). Asthma in Philadelphia schools. *Chest*, **124** (4): 141S.
113. Mangione, S., Yuen, E., and Madigan, D. (2003). Asthma and tobacco: A survey of 65 Philadelphia middle schools. *Chest*, **124** (4): 141S-142S.
114. Dunbar, P.J., Madigan, D., Grohskopf, L.A., Revere, D., Woodward, J., Minstrell, J., Frick, P.A., Simoni, J.M., and Hooton, T.M. (2003). A two-way messaging system to enhance antiretroviral adherence. *Journal of The American Medical Informatics Association*, **10**, 11-15.
115. Ridgeway, G. and Madigan, D. (2002). Bayesian analysis of massive datasets via particle filters In *Proceedings of KDD-02, The Eighth International Conference on Knowledge Discovery and Data Mining*, 5-13.
116. Madigan, D., Raghavan, N., DuMouchel, W., Nason, M., Posse, C., and Ridgeway, G. (2002). Likelihood-based data squashing: A modeling approach to instance construction. *Journal of Data Mining and Knowledge Discovery*, **6**, 173-190.
117. Tanimoto, S., Carlson, A., Husted, J., Hunt, E.B., Larsson, J., Madigan, D., and Minstrell, J. (2002). Text Forum Features for Small Group Discussions with Facet-Based Pedagogy. *Proceedings of CSCCL2002, Computer Supported Cooperative Learning*.
118. Hoeting, J., Raftery, A.E., and Madigan, D. (2002). A method for simultaneous variable and transformation selection in linear regression. *Journal of Computational and Graphical Statistics*, **11**, 485-507.
119. Liu, R., Madigan, D., and Eyheramendy, S. (2002). Text classification for mining aviation inspection reports. In: *Statistical Data Analysis based on the L1-norm and Related methods*. Birkhauser Statistics for Industry and technology, Y. Dodge editor, 379-392.

120. da Silva, C.Q., Zeh, J., Madigan, D., Laake, J., Rugh, D., Baraff, L., Koski, W., and Miller, G. (2001). Capture-recapture estimation of bowhead whale population size estimation using photo-identification data. *Journal of Cetacean Research and Management*, **2**, 45-61.
121. Levitz, M., Perlman, M.D., and Madigan, D. (2001). Separation and Completeness Properties for AMP Chain Graph Markov Models. *Annals of Statistics*, **29**, 1751–1784.
122. Church, L., Oliver, L., Dobie, S., Madigan, D., and Ellsworth, A. (2001). Analgesia for colposcopy: A double-blind, randomized comparison of ibuprofen and benzocaine gel for colposcopic analgesia. *Obstetrics and Gynecology*, **97**, 5-10.
123. Andersson, S.A., Madigan, D., and Perlman, M.D. (2001). An alternative Markov property for chain graphs. *Scandinavian Journal of Statistics*, **28**, 33-85.
124. Glusker, A.I., Dobie, S.A., Madigan, D., Rosenblatt, R.A., Larson, E.H. (2000). Differences in fertility patterns between urban and rural women in Washington state, 1983-1984 to 1993-1994. *Women and Health*, **31**, 55-70.
125. Kanungo, T., Haralick, R. M., Baird, H. Stuetzle, W., and Madigan, D. (2000). A statistical, nonparametric methodology for document degradation models validation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**, 1209-1223.
126. Dobie, S.A., Hart, G., Glusker, A., Rosenblatt, R., and Madigan, D. (2000). Reproductive health services in rural Washington state: Scope of practice and the potential of medical abortions. *American Journal of Public Health*, **90**, 624-626.
127. Madigan, D. and Nason, M. (2000). Statistics perspectives on data and knowledge. *Handbook of Knowledge Discovery and Data Mining*, Oxford University Press.
128. Nason, M. and Madigan, D. (2000). Sampling. *Handbook of Knowledge Discovery and Data Mining*, Oxford University Press.
129. Madigan, D., Raghavan, N., DuMouchel, W., Nason, M., Posse, C., and Ridgeway, G. (2000). Instance construction via likelihood-based data squashing. *Instance Selection and Construction, A Data Mining Perspective*, H. Motoda and H. Liu (Eds.), Kluwer, 209-226.
130. Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky, C.T. (1999). Bayesian model averaging – a tutorial. *Statistical Science*, **14**, 382-401.
131. Dobie, S.A., Hart, G., Glusker, A., Madigan, D., Larsen, E.B., and Rosenblatt, R. (1999). Abortion services in rural Washington State, 1983-1984 to 1993-1994: availability and outcomes. *Family Planning Perspectives*, **31**, 241-245.
132. Condliff, M.K., Lewis, D.D., Madigan, D., and Posse, C. (1999). Bayesian mixed-effects models for recommender systems. *Proceedings of SIGIR-99 Workshop on Recommender Systems*.
133. Madigan, D. (1999). Bayesian Graphical Models, Intention-to-Treat, and the Rubin Causal Model. In *Proceedings of Uncertainty-99, The Seventh International Workshop on Artificial Intelligence and Statistics*, 123-132.
134. Golinelli, D., Madigan, D., and Consonni, G. (1999). Relaxing the local independence assumption for quantitative learning in acyclic directed graphical models through hierarchical partition models. In *Proceedings of Uncertainty-99, The Seventh International Workshop on Artificial Intelligence and Statistics*, 203-208.

135. Ridgeway, G., Madigan, D., and Richardson, T. (1999). Boosting Methodology for Regression Problems. In *Proceedings of Uncertainty-99, The Seventh International Workshop on Artificial Intelligence and Statistics*, 152-161.
136. Ridgeway, G., Madigan, D., Richardson, T., and O’Kane, K. (1998). Interpretable Boosted Naïve Bayes Classification. In *Proceedings of KDD-98, The Fourth International Conference on Knowledge Discovery and Data Mining*, 101-104.
137. Andersson, S.A., Madigan, D., Perlman, M.D., and Richardson, T. (1998). Graphical Markov Models in multivariate analysis. In *Multivariate Analysis, Design of Experiments, and Survey Sampling*, Subir Ghosh (Ed.), Marcel Dekker Inc.
138. Draper, D. and Madigan, D. (1997). The scientific value of Bayesian statistical methods. *IEEE Intelligent Systems and their Applications*, **12**, 18-21.
139. Madigan, D. and York, J. (1997). Bayesian methods for estimating the size of a closed population. *Biometrika*, **84**, 19-31.
140. Glymour, C., Madigan, D., Pregibon, D., and Smyth, P. (1997). Statistical themes and lessons for data mining. *Journal of Data Mining and Knowledge Discovery*, **1**, 11-28.
141. Volinsky, C.T., Madigan, D., Raftery, A.E., and Kronmal, R.A. (1997). Bayesian Model Averaging in Proportional Hazard Models: Predicting Strokes. *Applied Statistics* **46**, 433-448.
142. Andersson, S.A., Madigan, D., Perlman, M.D., and Triggs, C.M. (1997). A graphical characterization of lattice conditional independence models. *Annals of Mathematics and Artificial Intelligence*, **21**, 27-50.
143. Madigan, D., Mosurski, K., and Almond, R.G. (1997). Explanation in belief networks. *Journal of Computational and Graphical Statistics*, **6**, 160-181.
144. Andersson, S.A., Madigan, D., and Perlman, M.D., (1997). A characterization of Markov equivalence classes for acyclic digraphs. *Annals of Statistics*, **25**, 505-541.
145. Raftery, A.E., Madigan, D., and Hoeting, J. (1997). Accounting for model uncertainty in linear regression. *Journal of the American Statistical Association*, **92**, 179-191.
146. Madigan, D., Keim, M., and Lewis, D.D. (1997). Bayesian information retrieval. *Proceedings of the Sixth International Workshop on Artificial Intelligence and Statistics*, 303-310.
147. Andersson, S.A., Madigan, D., and Perlman, M.D. (1997). On the Markov equivalence of chain graphs, undirected graphs, and acyclic digraphs. *Scandinavian Journal of Statistics*, **24**, 81-102.
148. Madigan, D., Raftery, A.E., Volinsky, C.T., and Hoeting, J.A. (1996). Bayesian model averaging. In: *Integrating Multiple Learned Models (IMLM-96)*, P. Chan, S. Stolfo, and D. Wolpert (Eds.), 77-83.
149. Raftery, A.E., Madigan, D., and Volinsky, C.T. (1996). Accounting for model uncertainty in survival analysis improves predictive performance. In: Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith A. F. M., (eds.), *Bayesian Statistics V*, Oxford University Press, 323-350.
150. Andersson, S.A., Madigan, D., and Perlman, M.D. (1996). An alternative Markov property for chain graphs. In *Proceedings of the Twelfth Annual Conference on Uncertainty in Artificial Intelligence*, Eric Horvitz and Finn Jensen (Eds.), Morgan Kaufmann Publishers, Inc., San Mateo.
151. Schaffner, A., Madigan, D., Hunt, E., Graf, E., Minstrell, J., and Nason, M. (1996). Benchmark lessons and the world wide web: Tools for teaching statistics. In: D. Edelson and E. Domeshek (Eds.) *Proceedings of ICLS 96. Association for the Advancement of Computing in Education*, 480-484.

152. Madigan, D. and Almond, R.G. (1996). On test selection strategies for belief networks. In *Learning from Data: Artificial Intelligence and Statistics V*, D. D. Fisher and H. Lenz (Eds.), Springer Verlag, 89-98 (20% acceptance rate for plenary papers).
153. Glymour, C., Madigan, D., Pregibon, D., and Smyth, P., and (1996). Statistical inference and data mining. *Communications of the ACM*, **39**, 35-41.
154. Madigan, D., Andersson, S.A., Perlman, M.D, and Volinsky, C.T. (1996). Bayesian model averaging and model selection for Markov equivalence classes of acyclic digraphs. *Communications in Statistics - Theory and Methods*, **25**, 2493-2519.
155. Hoeting, J.A., Madigan, D., and Raftery, A.E. (1996). A Method for Simultaneous Variable Selection and Outlier Identification in Linear Regression, *Journal of Computational Statistics and Data Analysis*, **22**, 251-270.
156. Madigan, D., Chapman, C.R., Gavrin, J., Villumsen, O., and Boose, J.H. (1995). Repertory hypergrids for large-scale hypermedia linking. *International Journal of Human-Computer Studies*, **43**, 465-481.
157. Madigan, D., Gavrin, J., and Raftery, A.E. (1995). Eliciting prior information to enhance the predictive performance of Bayesian graphical models. *Communications in Statistics - Theory and Methods*, **24**, 2271-2292.
158. Madigan, D. and York, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review*, **63**, 215-232.
159. Andersson, S.A., Madigan, D., Perlman, M.D., and Triggs, C.M., (1995). On the relation between conditional independence models determined by finite distributive lattices and by directed acyclic graphs. *Journal of Statistical Planning and Inference*, **46**, 25-46.
160. York, J., Madigan, D., Heuch, I., and Lie, R.T. (1995). Estimation of the proportion of congenital malformations using double sampling: Incorporating covariates and accounting for model uncertainty. *Applied Statistics*, **44**, 227-242.
161. Madigan, D., and Chapman, C.R. (1995). Multimedia tools for cancer pain education. In: *Medical Multimedia*, C. Ghaoui and R. Rada (Eds.), Intellect, Oxford, 121-136.
162. Haynor, D. and Madigan, D. (1995). Bayesian approach to differential diagnosis with use of Monte Carlo technique. *Radiology*, **197**, 425S.
163. Hanks, S., Madigan, D., and Gavrin, J. (1995). Temporal reasoning in probabilistic knowledge-based systems. In *Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence*, Philippe Besnard and Steve Hanks (Eds.), Morgan Kaufmann Publishers, Inc., San Mateo, 111-119.
164. Madigan, D., Chapman, C.R., Gavrin, J., Villumsen, O., and Boose, J.H. (1994). Repertory Hypergrids: An Application to Clinical Practice Guidelines. *Proceedings of the ACM European Conference on Hypermedia Technology*, 117-125.
165. Almond, R, Bradshaw, J.M., and Madigan, D. (1994). Reuse and sharing of graphical belief network components. In *Selecting Models from Data: Artificial Intelligence and Statistics IV*, P. Cheeseman and W. Oldford (Eds.), Springer Verlag, 113-122.
166. Madigan, D., York, J.C., Bradshaw, J.M., and Almond, R.G. (1994). Bayesian graphical models for predicting errors in databases. In *Selecting Models from Data: Artificial Intelligence and Statistics IV*, P. Cheeseman and W. Oldford (Eds.), Springer Verlag, 123-132.

167. Madigan, D., Raftery, A.E., York, J.C., Bradshaw, J.M., and Almond, R.G. (1994). Strategies for graphical model selection. In *Selecting Models from Data: Artificial Intelligence and Statistics IV*, P. Cheeseman and W. Oldford (Eds.), Springer Verlag, 91-100.
168. York, J.C. and Madigan, D. (1994). Markov chain Monte Carlo methods for hierarchical Bayesian expert systems. In *Selecting Models from Data: Artificial Intelligence and Statistics IV*, P. Cheeseman and W. Oldford (Eds.), Springer Verlag.
169. Madigan, D., Chapman, C.R., Gavrin, J., Villumsen, O., and Boose, J.H. (1994). Modularized Maintenance for Hyperlinking: An Application to Clinical Practice Guidelines. *Proceedings of the AAAI-94 Workshop on Indexing and Reuse in Multimedia Systems*, 126–140.
170. Kanungo, T., Haralick, R. M., Stuetzle, W., and Madigan, D. (1994). Document Degradation Models: Parameter Estimation and model validation. In *Proc. of Int. Workshop on Machine Vision Applications*, Kawasaki, Japan, 552-557.
171. Madigan, D. and Raftery, A.E. (1994). Model Selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association*, **89**, 1535–1546.
172. Madigan, D. (1993). A note on equivalence classes of directed acyclic independence graphs. *Probability in the Engineering and Informational Sciences*, **7**: 409–412.
173. Bradshaw, J.M., Chapman, C.R., Sullivan, K.M., Almond, R.G., Madigan, D., Zarley, D., Gavrin, J., Nims, J., and Bush, N. (1993). KS-3000: Applying DDUCKS to bone-marrow transplant patient support. *Proceedings of the Sixth Annual Florida AI Research Symposium (FLAIRS '93)*, Ft. Lauderdale, FL, April, 78–83.
174. Bradshaw, J.M., Madigan, D., Richards, T., and Boy, G.A. (1993). Emerging technology and concepts for computer-based training. *Proceedings of the Sixth Annual Florida AI Research Symposium (FLAIRS '93)*, Ft. Lauderdale, FL, April, 89–95.
175. Bradshaw, J.M., Richards, T., Fairweather, P., Buchanan, C., Guay, R., Madigan, D., and Boy, G.A. (1993). New directions in computer-based training in aerospace. *Proceedings of the Fourth International Conference on Human Machine Interactions and Artificial Intelligence in Aerospace*. Toulouse, France.
176. Bradshaw, J.M., Chapman, C.R., Sullivan, K.M., Boose, J.H., Almond, R.G., Madigan, D., Zarley, D., Gavrin, J., Nims, J. and Bush, N. (1993). KS-3000: An application of DDUCKS to bone-marrow transplant patient support. *Proceedings of the Seventh European Knowledge Acquisition for Knowledge-Based Systems Workshop (EKAW-93)*. Toulouse and Caylus, France, 57-74.
177. E.A. Kiely, Madigan, D., P.C. Ryan and M.R. Butler (1991). Ultrasonic imaging for extracorporeal shockwave lithotripsy: Analysis of factors in successful treatment. *British Journal of Urology*, **66**: 127–131.
178. Madigan, D. and Mosurski, K. (1991). An extension of the results of Asmussen and Edwards on collapsibility in contingency tables. *Biometrika*, **77**: 315–319.

Letters to the Editor/Commentaries

179. Madigan, D. (2021). Supra-disciplinary data science. *Harvard Data Science Review*, Issue 3.2.
180. Egilman, D., Madigan, D., Yimam, M., and Tran, T. (2020). Evidence that cosmetic talc is a cause of ovarian cancer. *Gynecology and Pelvic Medicine*, <http://dx.doi.org/10.21037/gpm-20-28>.

181. Egilman, D., Madigan, D., Yimam, M., and Tran, T. (2019). Letter to the Editor: Response to Vermont Talc-Miners Cohort Study Update. *Journal of Occupational and Environmental Medicine*, PMID: 31790061. DOI: 10.1097/JOM.0000000000001783.
182. Madigan, D., Egilman, D., Finkelstein, M., Yimam, M., and Tran, T. (2019). Response to Marsh, G. M., Ierardi, A. M., Benson, S. M., & Finley, B. L. (2019). Occupational exposures to cosmetic talc and risk of mesothelioma: an updated pooled cohort and statistical power analysis with consideration of latency period. *Inhalation Toxicology*, 31(6), 213–223.
183. Schuemie, M., Ryan, P., Hripcsak, G., Madigan, D., and Berlin, J. and Reich, C. (2019). Comment on "How pharmacoepidemiology networks can manage distributed analyses to improve replicability and transparency and minimize bias" *Pharmacoepidemiology and Drug Safety*, DOI: 10.1002/pds.4798
184. Schuemie, M., Ryan, P., Hripcsak, G., Madigan, D., and Suchard, M. (2016). Robust empirical calibration of p-values using observational data: Comment on 'Limitations of empirical calibration of p-values using observational data. *Statistics in Medicine*, DOI: 10.1002/sim.6977.
185. Schuemie, M.J., Ryan, P.B., Suchard, M., Shahn, Z., and Madigan, D. (2013). Discussion of “An estimate of the science-wise false discovery rate and application to the top medical literature” by Jager and Leek, *Biostatistics*.
186. Caster, O., Noren, N., Madigan, D., and Bate, A. (2013). Logistic regression in signal detection: another piece added to the puzzle (letter). *Clinical Pharmacology & Therapeutics*, 94, 312.
187. Ryan PB, Madigan D, Stang PE, Marc Overhage J, Racoosin JA, Hartzema AG. (2013) Response to Comment on 'Empirical assessment of methods for risk identification in healthcare data'. *Statistics in Medicine*, Mar 15;32(6):1075-7. doi: 10.1002/sim.5725.
188. Maclure, M., Fireman, B., Nelson, J., & Madigan, D. (2012). Mortality and the self-controlled case series method. Response to Letter to Editor. *Pharmacoepidemiology and Drug Safety*, 21(8), 907-907.
189. Madigan, D. and Gelman, A. (2009). Comment on “What is Statistics,” *American Statistician*, **63**, 114.
190. Bilker, W., Gogolak, V., Goldsmith, D., Hauben, M., Herrera, G., Hochberg, A., Jolley, S., Kulldorff, M., Madigan, D., Nelson, R., Shapiro, A., Shmueli, G. (2006). Accelerating Statistical Research in Drug Safety, Letter to the Editor, *Pharmacoepidemiology and Drug Safety*, **15**, 687-688.
191. Madigan, D. and Stuetzle, W. (2004). Statistics graduate education. Discussion of “The future of statistics” *Statistical Science*, **19**, 3, 408.
192. Madigan, D. and Ridgeway, G. (2003). Discussion of “Least Angle Regression” by Efron, Hastie, Johnstone, and Tibshirani. *Annals of Statistics*.
193. Madigan, D. and Ridgeway, G. (2002). Discussion of “Chain graph models and their causal interpretations” by Lauritzen and Richardson. *Journal of the Royal Statistical Society (Series B)*.
194. Madigan, D. (1999). Discussion of “Bayesian data mining in large frequency tables” by Bill DuMouchel. *The American Statistician*, **53**, 198-200.
195. Richardson, T., Ridgeway, G., and Madigan, D. (1999). Discussion of Bump Hunting in High-Dimensional Data by Jerome Friedman, *Statistics and Computing*, **9**, 150-152.
196. Higdon, D. and Madigan, D. (1998). Discussion of paper by Lavine in *Bayesian Statistics, VI*, 381-382.

197. York, J. and Madigan, D. (1993). Discussion of the paper by Smith and Roberts, *Journal of the Royal Statistical Society (Series B)*, **55**, 88.
198. Madigan, D. (1993). What's next? Contribution to *Statistical Science* discussion of two papers on graphical models, **8**:261–263.
199. Madigan, D. (1995). Discussion of the paper by Draper *Journal of the Royal Statistical Society (Series B)*, **57**, 85..
200. Madigan, D. (1995). Discussion of the paper by Chatfield *Journal of the Royal Statistical Society (Series A)*, **158**, 458-459.

Other Publications

201. Egilman, D., Madigan, D., Yimam, M., and Tran, T. (2020). Evidence that cosmetic talc is a cause of ovarian cancer. *Gynecology and Pelvic Medicine*. doi: 10.21037/gpm-20-28
202. Mulgrave, J.J., Madigan, D., and Hripcsak, G. (2020). Bayesian Posterior Interval Calibration to Improve the Interpretability of Observational Studies. arXiv:2003.06002, 2020
203. Gelman, A. & Madigan, D. (2015). Ethics and Statistics: How is Ethics Like Logistic Regression? Ethics decisions, like statistical inferences, are informative only if they're not too easy or too hard. *CHANCE*, 28(2), 31-33.
204. Ryan, P., Madigan, D., and Schuemie, M. (2014). The Emerging Role of Observational Healthcare Data in Pharmacovigilance. In: *Quantitative Evaluation of Safety in Drug Development: Design, Analysis and Reporting* edited by Qi Jiang and Amy Xia.
205. Stang, P., Patrick Ryan, Abraham G. Hartzema, David Madigan, J Marc Overhage, Emily Welebob, Christian G. Reich, Thomas Scarnecchia (2014). Development and Evaluation of Infrastructure and Analytic Methods for Systematic Drug Safety Surveillance: Lessons and Resources from the Observational Medical Outcomes Partnership. In: *Mann's Pharmacovigilance*, 3rd Edition Elizabeth B. Andrews and Nicholas Moore (Eds), Chapter 28.
206. McCormick, T.H., Rudin, C. and Madigan, D. (2011). Predicting medical conditions with Bayesian hierarchical rule modeling. Proceedings of the 6th INFORMS Workshop on Data Mining and Health Informatics (DM-HI 2011. P. Qian, Y. Zhou, C. Rudin, eds.
207. Hauben, M., Madigan, D., Patadia, V., Sakaguchi, M., van Puijenbroek, E. (2010). Quantitative signal detection for vaccines. *Human Vaccines*, 6, 1.
208. Madigan, D. (2007). Introduction to the LARS chapter. Volume celebrating Efron's 60th birthday, Springer.
209. Dayanik, A., Genkin, A., Kantor, P., Lewis, D.D., and Madigan, D. (2005). DIMACS at the TREC 2005 Genomics Track. TREC 2005.
210. Eyheramendy, S., Genkin, A., Ju, W-H., Lewis, D.D., and Madigan, D. (2003). Sparse Bayesian classifiers for text categorization. *JICRD*.
211. Madigan, D. (2003). Sparse Bayesian classifiers for text categorization. *Proceedings of the International Statistical Institute*.
212. Chaudhuri, S., Madigan, D., and Fayyad, U.M. (2000). KDD-99: The Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. *SIGKDD Explorations*, 1, 49-51.

213. Nakamura, Y., Chabal, C., Chapman, C.R., Dunbar, P.J., Madigan, D., and Minstrell, J. (1997). FABLE: A computer-based tool for teaching geriatric pain management skills. In *Proceedings of the Annual Conference of the American Pain Society*.
214. Madigan, D., Perlman, M.D., and Volinsky, C.T. (1995). Bayesian model averaging and model selection for Markov equivalence classes of acyclic digraphs. *Proceedings of Workshop on Model Robustness and Model Uncertainty*, Bath, England,
<http://www.isds.duke.edu:80/conferences/bath/abstracts.html>.
215. Madigan, D. (1995). Editorial for special issue of the *International Journal of Human-Computer Studies* on Knowledge-Based Hypermedia, **43**, 279.
216. Clarkson, D.B., Donnell, D., Minstrell, J., Hunt, E., Madigan, D., and Traynor, C. (1994). Vital: An intelligent tutoring system for statistics. *American Statistical Association, Proceedings of Section on Statistical Education*, 88-93.
217. Dunbar, P.J., Madigan, D., Lam, A.M., and Matta, B.F. (1994). A hypermedia instruction tool for teaching retrograde jugular venous cannulation. Multimedia Scientific Exhibit at the 1994 Meeting of the American Society of Anesthesiologists, San Francisco, CA.
218. Madigan, D. (1989). Microcomputer research at the Meath hospital. *Irish Medical Times*, **21** (34), 14-15.
219. Madigan, D. (1989). Expert systems in government. *Irish Computer*, 22-24.

Book and Software Reviews

220. Madigan, D. (2002). A review of "Principles of Data Mining" by Hand, Mannila, and Smyth. *SIAM Review*, **44**, 501-502.
221. Madigan, D. (2001). A review of "Probabilistic Networks and Expert Systems" by Cowell, Dawid, Lauritzen, and Spiegelhalter. *Journal of the American Statistical Association*, **96**, 1524.
222. Madigan, D. (1994). A review of MIM: graphical modelling software. *Statistics and Computing*, **4**, 33-39.
223. Madigan, D. (1994). A review of "Graphical models in applied multivariate statistics" by Joe Whittaker. *Networks*, **24**, 125.

Unpublished Technical Reports

224. Brookhart, M.A., Ryan, P., Madigan, D., Sturmer, T. (2011). An Empirical Comparison of Different Implementations of a Standardized New User Design For Drug Safety Surveillance.
225. Egilman, D., Madigan, D., and Drugar, N.M. (2011). A Drug Trial Gone Wrong: Excess Death and Injury Among Study Volunteers in an Unmonitored Alzheimer's Drug Study.
226. Ryan, P.B., Reich, C., Welebob, E., Overhage, J.M., Stang, P.E., Hartzema, A.G., Racoosin, J.A., Scarnecchia, T., Madigan, D. (2011). Managing data quality for an active surveillance system.
227. Pickering, W.H., Madigan, D., McCarter, R.J., and Burd, R.S. (2009). Evaluating relative importance of injury groupings on in-hospital mortality.
228. Hauben, M., Madigan, D., Reisinger, S., Hochberg, A., and O'Hara, D. (2008). Effects of Stratification on Three Pharmacovigilance Data Mining Algorithms.

- 229. Ju, Wen-Hua, Madigan, David, and Scott, Steven (2002). On Bayesian learning of sparse classifiers.
- 230. J.W. O’Kane, G. Ridgeway, and D. Madigan (1999). Statistical Analysis of Clinical Variables to Predict the Outcome of Surgical Intervention in Patients with Knee Complaints.
- 231. Madigan, D. (1998). Combining probability distributions: A Review. Statistical Sciences Inc. Research Report.
- 232. Schaffner, A., Madigan, D., Clarkson, D.B., Donnell, D., Hunt, E.B., Keim, M., Minstrell, J., Nason, M., and Volinsky, C.T. (1996). Facet-based learning for statistics.
- 233. Schaffner, A., Madigan, D., Hunt, E.B., and Minstrell, J. (1996). Virtual benchmark instruction.
- 234. Madigan, D., Hunt, E., Levidow, B., and Donnell, D. (1995). Bayesian graphical models for intelligent tutoring systems.
- 235. Madigan, D. (1992). Temporal Reasoning with Probabilities: A Review. Statistical Sciences Inc. Research Report Number 7.
- 236. Madigan, D. (1992). Approaches to Explanation in Bayesian Networks. Statistical Sciences Inc. Research Report Number 8.
- 237. Carlsen, J.C., Madigan, D. and Bradshaw, D. (1992). Music expectancy and its measurement. UW Department of Music Technical Report.

Research Grants

Principal Investigator on sub-contract to Northeastern University from FDA Award 75F40120D0039 to Columbia University, 2020-2021, \$543,138.

Principal Investigator on sub-contract from NSF Award IIS 1251151 to UCLA, “Patient-level predictive modeling from massive longitudinal databases.” 2013-2017, \$217,837.

Principal Investigator on FNIH/Observational Medical Outcomes Partnership grant to Columbia University, “Methods for Active Drug Safety Surveillance,” 2009-2013.

Co-Principal Investigator on National Institutes of Health grant “Improving Pediatric Trauma Triage Using High Dimensional Data Analysis,” R01 GM87600-01, 2010-2013.

Principal Investigator on sub-contract from FDA Award HHSF223200910006I to Harvard Pilgrim Healthcare, “Mini-sentinel,” 2010-2011.

Investigator on Department of Homeland Security grant “Center for Dynamic Data Mining” to Center for Discrete Mathematics and Theoretical Computer Science (DIMACS), Rutgers University, 2006-2009, \$2,400,000 total.

Principal Investigator on National Science Foundation grant “Bayesian Methods for Large-Scale Applications” to Columbia University, DMS-0505599, 2005-2009, \$150,000 total.

Principal Investigator (with Shlomo Argamon) on National Science Foundation grant “Community resources for author identification” to Rutgers University, CNS-0454126, 2005, \$56,019 total.

Principal Investigator on Knowledge Discovery and Dissemination (KD-D) grant to DIMACS “Author Identification,” 2004, \$250,000 total.

Co-Principal Investigator on National Science Foundation grant “Monitoring Message Streams” to DIMACS, Rutgers University, KDI-0087022, 2002-2008, \$1,500,000 total (annual renewal).

Investigator on National Science Foundation grant “Computational and Mathematical Epidemiology” to DIMACS, Rutgers University, ITR-0205116, 2002-2007, \$2,750,000 total.

Principal Investigator on National Science Foundation grant “Bayesian Data Analysis for Digital Traces” to Rutgers University, DMS-0113236, 2001-2004, \$245,000 total.

Principal Investigator (with Steen Andersson and Michael Perlman) on National Science Foundation grant “Graphical Markov Models” to the University of Washington, DMS-9704573, 1997-2000, \$313,722 total.

Principal Investigator (with Steven Tanimoto) on grant from National Science Foundation to University of Washington: “Use of Online Assessment in Forming and Coaching Learning Groups,” 1996-1999, \$600,501, CRLT-9616532.

Principal Investigator on National Science Foundation grant “Computing Environments for Graphical Models” to the University of Washington, DMS-92111629, 1992-1996, \$65,000 total.

Principal Investigator on Subcontract to the University of Washington under National Institutes for Health Phase II SBIR grant “TALARIA: Multimedia tools for cancer pain education” to Mathsoft, Inc., 1995-1997, \$750,000 total (UW Subcontract, \$231,652).

Principal Investigator on Subcontract to the University of Washington under National Institutes for Health Phase I SBIR grant “An intelligent tutoring system for biostatistics” to Statistical Sciences, Inc., 1994-5, \$80,853.

Principal Investigator on National Institutes for Health SBIR Phase I grant “A Multimedia Teaching and Reference Tool for Cancer Pain” to Statistical Sciences, Inc., 1993-1994, \$50,000 total.

Investigator (20%) on Quantum Health Resources Grant “Knowledge-based Systems for Bone Marrow Transplant Long Term Follow-up” to the Fred Hutchinson Cancer Research Center, 1993-94, \$1m (C. Richard Chapman, P.I.).

Investigator (and author) on Department of Energy Phase I SBIR grant “An intelligent tutoring system for statistics” to Statistical Sciences, Inc., 1993-4, \$75,000.

Investigator (10%) on Robert Wood Johnson Foundation grant to the Department of Family Medicine, University of Washington (Sharon Dobie, PI), 1995-1999.

Investigator (5%) on American College of Clinical Pharmacy grant to the Department of Family Medicine, University of Washington (Allan Ellsworth, PI), 1995-1996.

Patents

Ju, W., Krishnakumar, A.S., Krishnan, P., and Madigan, D. (2008). Method and apparatus for positioning a set of terminals in an indoor wireless environment. US PTO # 7403784.

Selected Invited Presentations

ISPOR 2021 (virtual), “We have the technology,” April 2021.

Temple University, Data Science Meeting (virtual), “Real Real-World Evidence” November 2020.

Joint Statistical Meeting (virtual), “OHDSI Methods for Causal Inference” August 2020.

Advances in Precision Medicine Conference, Columbia University (virtual), “Real Real-World Evidence” April 2020.

EPFL, Switzerland (virtual), “Real Real-World Evidence” April 2020.

University at Buffalo, “Real Real-World Evidence” November 2019.

AMS Regionakl Meeting, Riverside, CA, “Real Real-World Evidence,” November 2019.

ASA NJ Chapter / Bayer 7th Annual Workshop, Keynote Speaker, “Real Real-World Evidence,” November 2019.

National Academy of Science, Applying Big Data to Address the Social Determinants of Health in Oncology, “Real Real-World Evidence,” October 2019.

University of Florida Informatics Institute, Annual Symposium, Keynote, “Real Real-World Evidence,” October 2019.

Harvard University Department of Healthcase Policy, “Real Real-World Evidence” September 2019.

U. Penn Big Data in Healthcare Meeting, Washington, DC, “Real-World Evidence” September 2019.

Duke-Margolis Meeting on Leveraging RCTs to Generate Real-World Evidence, Washington, DC, “Real-World Outcomes” July 2019.

NYC R Conference, New York, “Honest Inference from Observational Studies” May 2019.

Data Science Dean’s Speaker Series, SUNY Binghamton, “Honest Inference from Observational Studies” April 2019.

Neyman Lecture, University of California, Berkeley, “Honest Inference from Observational Studies” March 2019.

Arizona State University, Phoenix, “Honest Inference from Observational Studies” November 2018.

Boston University, Boston, “Ethical Challenges in Drug Development” October 2018.

IQVIA Research Institute Forum, Boston, “Honest Inference from Observational Studies” July 2018.

Keynote Address, FDA Center for Biologics Evaluation Research Science Day, “Honest Inference from Observational Studies” June 2018.

ASA Conference on Statistical Learning and Data Science, Keynote Address, “Honest Inference from Observational Studies” June 2018.

Atlantic Causal Inference Conference, “A Bayesian approach to modeling negative controls in observational studies.” CMU, May 2018.

University of Arizona, “Honest Inference from Observational Studies” April 2018.

Temple University, “Honest Inference from Observational Studies” April 2018.

CRM Montreal, Workshop on Risk Modeling, Management and Mitigation in Health Sciences, “A data-driven world: opportunities and challenges,” December 2017.

Royal Society London, Workshop on the Ubiquity of Algorithms, “A data-driven world: opportunities and challenges,” October 2017.

New York University, Center for Data Science, “A data-driven world: opportunities and challenges,” October 2017.

Pfizer Analytics Summit, “Honest Inference from Observational Studies” October 2017.

American Express, New York, “A data-driven world: opportunities and challenges,” September 2017.

IBM Watson Computational Health Summit, “Honest Inference from Observational Studies” May 2017.

IMS Spring Research Conference, Rutgers, Keynote Address, “Honest Inference from Observational Studies” May 2017.

New England Statistics Symposium, U Conn, Keynote Address, “Honest Inference from Observational Studies” April 2017.

IBM Watson, Yorktown Heights, NY, “Honest Inference from Observational Studies” March 2017.

Sackler Meeting at the National Academy of Sciences, Washington, DC, “Honest Inference from Observational Studies” March 2017.

University of Wisconsin, Department of Biostatistics, “Honest Inference from Observational Studies” November 2016.

Vanderbilt University, Department of Biostatistics, “Honest Inference from Observational Studies” October 2016.

Dana-Farber Reproducibility in Personalized Medicine Research Workshop, Boston, “Honest Inference from Observational Studies,” September 2016.

EXPERT 2016: Trailblazers, “Large-scale Observational Healthcare Data: Promise and Peril,” IUPUI, Indianapolis, September 2016.

International Society for Pharmacoepidemiology and Therapeutic Risk Management, Annual Meeting, Dublin, Ireland, August 2016.

International Chinese Statistical Association Applied Statistics Meeting, Keynote Address, Atlanta, “Honest Inference from Observational Studies” June 2016.

Trends and Innovations in Clinical Trial Statistics, Keynote Address, North Carolina, “Honest Inference from Observational Studies” May 2016.

Penn State University, Department of Statistics, “Honest Inference from Observational Studies” April 2016.

Theory of Big Data Conference, London, “Honest Inference from Observational Studies” January 2016.

University of Michigan, Department of Statistics, “Honest Inference from Observational Studies” October 2015.

ASA Connecticut Chapter, Farmington, CT, “Honest Inference from Observational Studies” April 2015.

Sackler Meeting at the National Academy of Sciences, Washington, DC, “Honest Inference from Observational Studies” March 2015.

NISS Affiliates Workshop, Miami FL, “Observational Studies: Lessons from OMOP and OHDSI” March 2015.

ENAR, Miami FL, “Observational Studies: Lessons from OMOP and OHDSI” March 2015.

U. Mass Medical School Grand Rounds, Worcester, MA, “Observational Studies: Lessons from OMOP and OHDSI” February 2015.

11th Global Cardiovascular Clinical Trials Forum, Washington, DC, “Honest Inference from Observational Studies” December 2014.

Duke University, Department of Statistical Science, “Honest Inference from Observational Studies” October 2014.

Good Medical Research Conference, Cooper Union, New York, “Honest Inference from Observational Studies” October 2014.

First Seattle Symposium on Healthcare Data Analytics, Group Health, Seattle, “Honest Inference from Observational Studies” September 2014.

Institute for Data Sciences and Engineering, Columbia University, “Are Observational Studies Any Good?” September 2014.

Joint Statistical Meetings, Boston, Lunch with the Speaker, “Are Observational Studies Any Good?” August 2014.

Rutgers Annual Statistics Symposium, “Are Observational Studies Any Good?” May 2014.

Keynote Address, SIAM International Conference on Data Mining, Philadelphia, “Are Observational Studies Any Good?” April 2014.

Rustagi Lecture, Ohio State University, “Calibrating Observational Studies” April 2014.

ENAR, Baltimore, “Calibrating Observational Studies” March 2014.

Stern School, New York University, “Are Observational Studies Any Good?” March 2014.

PaSiPhIC Conference, San Luis Obispo, “Are Observational Studies Any Good?” Keynote address, February 2014.

Robert Wood Johnson/Rutgers, “Are Observational Studies Any Good?” December 2013.

Pfizer Inc., New York, “The Bayesian List Machine” October 2013.

University at Buffalo, “Are Observational Studies Any Good?” October 2013.

Joint Statistical Meetings, Montreal, “The Bayesian List Machine,” August 2013.

NYC R-Meetup, “Are Observational Studies Any Good?” July 2013.

IMA New Directions Summer School, University of Minnesota, June 2013.

McGill University, Department of Epidemiology, Biostatistics & Occupational Health, “Are Observational Studies Any Good?” February 2013.

Brown University, Department of Biostatistics, “Are Observational Studies Any Good?” November 2012.

Pfizer Inc., New York, “Are Observational Studies Any Good?” November 2012.

Yale University, Department of Biostatistics, “Are Observational Studies Any Good?” November 2012.

University of Montreal, Canada, “Are Observational Studies Any Good?” October 2012.

Rutgers University, Department of Statistics, “Are Observational Studies Any Good?” October 2012.

Carnegie Mellon University, Department of Statistics, “Are Observational Studies Any Good?” October 2012.

Yale University, Department of Statistics, “Are Observational Studies Any Good?” October 2012.

SAMSI Opening Workshop on Data-Driven Decisions in Healthcare, “A Predictivist Approach to Observational Analyses in Healthcare,” SAMSI, August 2012.

Atlantic Causal Inference Conference, “Big-Data-Driven Medicine,” Baltimore, May 2012.

Interface 2012, “Massive Parallelization of Serial Inference Algorithms for a Complex Generalized Linear Model,” Houston, TX, May 2012.

IMA Meeting on User-Centered Modeling, “Big-Data-Driven Medicine,” Minneapolis, May 2012.

New York City Center for Innovation Through Data Intelligence, “Big-Data-Driven Medicine,” New York City, April 2012.

University of Texas at Austin, “High-Dimensional Pharmacoepidemiology,” Austin TX, April 2012.

ENAR Spring Meeting, “High-Dimensional Pharmacoepidemiology,” Washington DC, April 2012.

Institute of Medicine Meeting on Healthcare Data, “High-Dimensional Pharmacoepidemiology,” Washington DC, March 2012.

Wharton Business School, “Statistical Methods for Drug Safety Surveillance: Big Data to the Rescue?” Philadelphia, December 2011.

New Paradigms in Clinical Trial Methodology Symposium, “Medicine meets Big Data,” Research Triangle Park, NC, November 2011.

Drug Information Association Annual Meeting, “OMOP – A Summary of the Findings,” Chicago, June 2011.

Drug Safety Research Unit 6th Biennial Conference - Signal Detection & Interpretation in Pharmacovigilance, London, “Signal Detection Methods,” June 2011.

New York Machine Learning Meet-up, New York, “Bayesian model averaging – new tricks for an old dog,” May 2011.

Café Science public lecture, New York, “How safe are your prescription drugs?” May 2011.

PhRMA/FDA Statistical Leaders Conference, Washington DC, “Bayesian methods in active surveillance,” April 2011.

AcademyHealth, Electronic Data Methods Forum Symposium, Washington, DC, “OMOP Initial Findings,” April 2011.

North Carolina State University, “Drug Safety”, April 2011

DIA Computational Science Meeting, Washington DC, “Self-controlled methods for analyzing recurrent events in large-scale longitudinal data,” March 2011.

Department of Mathematics and Statistics, Bowling Green State University, “Active Surveillance for Drug Safety,” Bowling Green, OH, December 2010.

Second Annual Princeton Day of Statistics, “Active Surveillance for Drug Safety,” Princeton, NJ, October 2010.

Workshop on Recent Advances in Bayesian Computation, “Big Bayesian Logistic Regression,” Singapore, September 2010.

Joint Statistical Meetings, “The Observational Medical Outcomes Partnership,” Vancouver, Canada, August 2010.

Sparsity Workshop, University of Bristol, “Sparse methods in drug safety,” Bristol, U.K., June 2010.

Valencia 9 International Meeting on Bayesian Statistics, “Bayesian methods in pharmacovigilance,” Benidorm, Spain, June 2010

CRISM Workshop on Model Uncertainty, “Sequential Bayesian Model Averaging,” Warwick, U.K., May 2010

Data Mining and Nonparametric Statistics conference, “Active Surveillance for Drug Safety,” Columbus, OH, May 2010.

International Society for Pharmacoepidemiology, mid-year meeting, “Observational Medical Outcomes Partnership: Methods Update,” Raleigh, NC, April 2010.

Department of Statistics, Harvard University, “Active Surveillance for Drug Safety,” Cambridge, MA, March 2010.

Frontiers of Statistical Decision Making and Bayesian Analysis (in honor of Jim Berger), “Active Surveillance for Drug Safety,” San Antonio, March, 2010.

American Society for Microbiology Biodefense Meeting, “A cross-species analysis of the CDC anthrax vaccine safety data,” Baltimore, February, 2010.

DIMACS 20th Anniversary Conference, “Drug safety, port security, and anthrax: A DIMACS medley.” New Jersey, November, 2009.

DIA 2nd Annual Conference on Signal Detection and Data Mining, “The OMOP Project.” New York, November 2009.

Keynote Speaker at 11th Annual Johnson & Johnson Statistics Conference, “Post-marketing drug safety surveillance: new developments,” New Jersey, October, 2009.

CDC Annual Anthrax Vaccine Research Meeting, “Correlates of protection, the bridge from animals to humans.” September 2009.

Joint Statistical Meetings, “Sequential Bayesian Model Selection,” Washington, DC, August 2009

IMS Medalion Lecture, WNAR Conference, “High Dimensional Bayesian Classifiers,” Portland, Oregon, June 2009.

Quality and Productivity Research Conference, “High Dimensional Bayesian Classifiers,” IBM Watson, June 2009.

Taft Competitive Lecture, University of Cincinnati, “How safe are your drugs?,” May, 2009

Department of Biomedical Informatics, Columbia University, “Shrinkage methods for drug safety,” New York, March 2009.

Roche Global Safety Science Meeting, “Logistic regression for drug safety,” Vienna, Austria, February 2009.

Department of Statistics, Rice University, “Shrinkage methods for drug safety,” Houston, TX, January 2009.

Department of Statistics, University of Illinois, “Shrinkage methods for drug safety,” Champaign, IL, December 2008.

Drug Information Association Signal Detection and Data Mining workshop, “Shrinkage methods for drug safety,” Washington DC, November 2008.

DIMACS Port Security Workshop, “Efficient sequential decision making algorithms for container inspection operations,” New Jersey, November 2008.

Psychiatry Institute, Columbia University, “Data mining and the drug development process: Safety,” New York, September 2008.

Institute of Mathematical Statistics Annual Meeting, “High Dimensional Bayesian Classifiers,” Singapore, July 2008.

International Conference on Machine Learning and Data Mining, half day tutorial, “Text Mining,” Beijing, China, June 2008.

International Conference on Machine Learning and Data Mining, “Data mining and the drug development process: Safety,” Beijing, China, June 2008.

International Indian Statistical Association Annual Meeting. “Data mining and the drug development process: Safety,” University of Connecticut, May 2008.

10th Annual Symposium on Statistics in Psychiatry, “High-dimensional Bayesian classifiers,” New York University, May 2008.

Brooklyn Law School, “Secrets of Vioxx: Lessons for Drug Safety,” April 2008.

Columbia University, Department of Applied Mathematics, “High-dimensional Bayesian classifiers,” April 2008.

NIH/NIAID, “Secrets of Vioxx: Lessons for Drug Safety,” March 2008.

CDC Anthrax Vaccine Correlate of Protection Meeting, “Predictive modeling building for correlates of protection,” March 2008.

Tenth Annual Winter Workshop, Bayesian Model Selection and Objective Methods, University of Florida, “High-Dimensional Bayesian Classifiers,” January 2008.

University of Pennsylvania Wharton Business School, “Secrets of Vioxx: Lessons for Drug Safety in the Drug Development Process,” December 2007.

University of Washington Department of Biostatistics, “Secrets of Vioxx: Lessons for Drug Safety in the Drug Development Process,” November 2007.

Annual meeting of the International Society of Pharmacovigilance, Bournemouth, UK, “How to Shrink in Pharmacovigilance.” October, 2007.

SIAM Conference on Mathematics for Industry, Philadelphia, PA, “Pharmacovigilance: new methods needed.” October, 2007.

U. Penn Invitational Choice symposium, Philadelphia, PA, “Statistical Analysis: Bigger and Bigger,” June 2007.

Midwest Biopharmaceutical Statistics Workshop, Muncie IN, “Bayesian post-marketing drug safety surveillance,” May 2007.

Yale University, “Lasso Logistic Regression: Recent Developments” April, 2007.

ENAR Annual Meeting, Atlanta GA, “Bayesian post-marketing drug safety surveillance,” March 2007.

Duke University, “Lasso Logistic Regression: Recent Developments” November, 2006.

CDC Anthrax Vaccine Research Program Seventh Annual Investigator’s Meeting, “Non-Human Primate Study,” Atlanta, GA, October 2006.

UCLA Undergraduate Statistics Program, “Localization in Wireless Networks”, two-day practicum, June 2006, Los Angeles.

Classification Society of North America Annual Meeting, “The Power of the Prior,” May 2006, New Jersey.

Workshop on The Science of Learning and the Teaching of Math and Science, “Facet-Based Learning,” May, 2006, Rutgers University.

Data Mining in Pharmacovigilance, April 2006, Pfizer, New York City

Data Mining in Pharmacovigilance, March 2006, DIA Meeting, Washington, DC

University of Chicago, “Sparse Bayesian Classification”, February 2006.

Data Mining in Drug Safety, February 2006, DIMACS Workshop

Bayesian Statistics VIII, June 2006, Valencia, Invited talk. (declined due to conflict)

Drug Information Association Tutorial on Data Mining for Pharmacovigilance, Washington D.C., January 2006.

CDC Anthrax Vaccine Research Program Sixth Annual Investigator’s Meeting, “Non-Human Primate Study: Interim Report,” Atlanta, GA, October 2005.

NISS Workshop in Honor of Jon Kettenring, Better Data Analysis with Prior Knowledge, September 2005.

SAMSI Workshop on Homeland Security and National Defence, Statistical Methods for Authorship Attribution, SAMSI, September 2005.

Mitre Workshop on the Significance of Bioinformatics to National Security, Washington D.C., June 2005

Plenary Address, Graybill Conference, “Text Mining,” Colorado, June 2005.

Google, New York, “Text mining,” April 2005

Princeton University, “Online logistic regression,” March 2005

American Statistical Association, Florida Chapter, Annual Meeting, “From Sewage to Guns: 20 Years of Statistical Consulting”, February 2005.

Florida State University, “The statistical analysis of text data”, February 2005

MCMSki Workshop, Bormio, Italy, “Text categorization”, January 2005.

M2004 SAS Data Mining Conference, “Text mining”, October, 2004

Pacific Northwest National Laboratory, “Bayesian Model Selection”, October, 2004

Bell Labs, “Bayesian Location Estimation”, October, 2004

JSM, Toronto, “Bayesian Location Estimation”, August, 2004

Bertinoro, Italy, Workshop on the Mathematics of Web Search, “Statistical Analysis of text Data”, June 2004.

Wharton Business School, “Bayesian text categorization,” April 2004.

DIMACS Workshop on Data Mining and Epidemiology, “Data Mining: An Overview”, March 2004.

Columbia University, “Statistical methods for the analysis of textual data”, September 2003.

University of Aalborg, Denmark, “Bayesian graphical models for location determination”, September 2003

SAMSI, “Statistical methods for the analysis of textual data”, September 2003.

ISI, Berlin, “Text Categorization”, August 2003.

DIMACS Working Group on Data Mining and Epidemiology, “Analysis of Hospital Discharge Data”, May 2003.

Johns Hopkins University, “Text categorization”, May 2003

North Carolina State University, “Text Categorization”, May 2003

Cleveland Clinic, “Graphical Markov Models”, December, 2002.

IBM TJ Watson, “Text Categorization”, November, 2002

Educational Testing Service, “Sequential Monte Carlo Methods for Massive Datasets”, November, 2002

University of Connecticut, “Text Categorization”, September, 2002

IMS Annual Meeting, Banff, Canada, “Retrieval properties of large collections,” July, 2002.

IMS Annual Meeting, Banff, Canada, “Bayesian analysis of hidden Markov models,” July, 2002.

Keynote Speaker, 22nd International Symposium on Forecasting, Dublin, “Bayesian analysis of hidden Markov models”, June 2002.

Bayesian Statistics VII, June 2002, Valencia, Invited discussion.

Duke University, “Text Categorization”, May 2002.

University of California, Irvine, “Text Categorization”, April 2002.

University of Southern California, “Text Categorization”, April 2002.

New York University, “Bayesian analysis of hidden Markov models”, April 2002.

DIMACS Epidemiology Workshop, “Some aspects of adverse events detection.” March 2002.

Temple University, “Text Categorization”, April 2002.

Haifa Winter Workshop on Computer Science and Statistics, Technion, Haifa, Israel, “Bayesian Analysis of Hidden Markov Models.” December, 2001.

Haifa Winter Workshop on Computer Science and Statistics, University of Haifa, Israel, “Likelihood-based Data Squashing.” December, 2001.

Columbia University, Graphical Models and Bayesian Networks: A History. October, 2001.

SCILS, Rutgers University, Bayesian statistical methods for digital traces, April, 2001.

Bell Labs, Data warehousing and reporting: a case study, January, 2001.

Rutgers University, Graphical Markov Models, December, 2000.

University of Washington. Data squashing. November 2000.

Royal Statistical Society Annual Meeting, University of Reading. Graphical Models and Bayesian Networks: A History. September, 2000.

University of Chicago Business School. Data squashing. November 1999.

Joint Statistical Meetings, August 1999, Baltimore. “Bayesian data mining in large frequency tables.”

UW Mathday Plenary Speaker. Data in-Garbage out: How to twist the truth with statistics. March 1999.

Bell Labs, Bayesian Model Selection. January 1999.

Seventh International Workshop on AI and Statistics, Florida. Bayesian Graphical Models, Intention-to-Treat, and the Rubin Causal Model. January, 1999.

Bayesian Statistics VI, June 1998, Valencia, Invited discussion.

Bellcore, "Bayesian Collaborative Filtering" June 1998.

Cambridge University, Statistical Laboratory, "Noncompliance in clinical trials," October 1997.

Isaac Newton Institute for Mathematical Sciences, Cambridge, UK "Bayesian model averaging," October 1997.

Trinity College Dublin, "Bayesian model averaging," October 1997.

UW/Microsoft Data Mining Institute, "Bayesian model averaging," July 1997.

KDD-97, "Graphical models," August 1997.

"The World-Wide Web as a Statistical Producer," Derry, Northern Ireland, "Statistical Analysis of Web-Generated Data," April 1997.

International Association for Statistical Computing, Pasadena, CA, "Graphical Markov Models for Chain Graphs," February 1997.

Sixth International Workshop on AI and Statistics, Florida, "Bayesian Information Retrieval," January, 1997.

Royal Statistical Society, Special Session sponsored by the Research Section, Invited Talk, "Dealing with model uncertainty," September, 1996.

ETS, Princeton, New Jersey, August 1996, "Bayesian model averaging."

Joint Statistical Meetings, August 1996, Chicago. "Bayesian information retrieval."

First European Conference on Highly Structured Stochastic Systems, Rebild, Denmark, May 1996.

INFORMS (formerly ORSA-TIMS) meeting, May 1996, Washington DC, "New developments in Bayesian model averaging."

Departments of Philosophy and Statistics, Carnegie-Mellon University, Invited talks, May 1996.

ENAR meetings, Richmond, VA, and ISDS, Duke University, "Model Selection and Averaging with Biostatistics applications," March 1996, invited.

Department of Mathematical Sciences, University of Alaska at Fairbanks, February 1996, invited.

NIPS Conference, Vail, Colorado December 1995, "Bayesian model averaging and model selection for Markov equivalence classes of acyclic digraphs"

Model Uncertainty Workshop, Bath, England, June 1995, "New developments in Bayesian model averaging."

Algebraic Methods in Multivariate Statistics, Oberwolfach, Germany, July 1995.

Workshop on Maximum Entropy and Bayesian Methods, Sante Fe, July 1995, "New developments in Bayesian model averaging."

Department of Psychology, UW, Facet-based Learning for Statistics, April, 1995

Fifth International Workshop on Artificial Intelligence and Statistics, Florida, Plenary Talk, “Test selection for graphical models,” January 1995.

Department of Statistics, UW, Facet-based Learning for Statistics, December, 1994

European Conference on Hypermedia Technology, Edinburgh, Scotland, September 1994, “Repertory hypergrids: An application to clinical practice guidelines.”

Bell Communications Research, August 1994, “Repertory hypergrids for hypermedia.”

Bell Laboratories, August 1994, “Model Uncertainty” .

IMS Meeting, June 1994, Chapel Hill, “Computations for Bayesian graphical models.”

Bayesian Statistics V, June 1994, Valencia, “Improving the predictive performance of Bayesian graphical models.”

The Seventh Annual Florida Artificial Intelligence Research Symposium, May 1994. “Building Bayesian models for intelligent tutoring systems.”

Joint Statistical Meetings, August 1993, San Francisco. “Accounting for Model Uncertainty.”

First ACM Workshop on Multimedia in Medical Education, August 1993, Anaheim. “Multimedia tools for Cancer Pain Education.”

Bayes Factors and Sensitivity Analysis Workshop, February 1993, UCLA. “Bayesian graphical models.”

Society for Medical Decision Making, October 1992, Portland, Oregon. “Bayesian Statistics.”

Joint Statistical Meetings, August, 1992, Boston. “Model Selection and Accounting for Model Uncertainty in Graphical Models using Occam’s Window.”

The Second International Conference on Music Perception and Cognition, February, 1992, Los Angeles. “Development of a Data-based Expectancy Model.”

Scholarly Service Activities

Scholarly Journals

Associate Editor, *Harvard Data Science Review*, 2018-.

Editorial Board, *Journal of Scientific Practice and Integrity*, 2018-

Member, Editorial Committee, *Annual Reviews of Statistics and its Application*, 2018-2021.

Editor-in-Chief, *Statistical Analysis and Data Mining – the ASA Data Science Journal*, 2013-2015.

Associate Editor, *Statistical Science*, 2011-2013.

Editor-in-Chief, *Statistical Science*, 2008-2010.

Editorial Board, *Therapeutic Innovation and Regulatory Science*, 2013-

Editorial Board, *International Journal of Occupational and Environmental Health*, 2012-2017

Advisory Board, Wiley Interscience Review Series (WIREs) On Data Mining And Knowledge Discovery, 2008-

Editorial Board, *Foundations and Trends in Machine Learning*, 2007-

Senior Associate Editor, *Advances in Disease Surveillance*, 2005-2009.

Advisory Board, *Bayesian Analysis*, 2004-2006.

Action Editor, *Journal of Machine Learning Research*, 2003-2006.

Associate Editor, *Journal of Computational and Graphical Statistics*, 1997-2002.

Associate Editor, *Journal of the Royal Statistical Society (Series B)*, 1995-1999.

Editor of special issue of the *International Journal of Human-Computer Studies*: “Knowledge-based hypermedia,” 1995.

Editorial Board for the *Handbook of Knowledge Discovery and Data Mining*, 1997-2000.

Editorial Board for the *Journal of Data Mining and Knowledge Discovery*, 1996-2004.

Professional Societies

Co-Chair, ACM-IMS Data Science Joint Venture, 2019-

Chair, ASA Breiman Award Committee, 2019-2020.

Member, International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force in Enhancing the Utility of Real World Evidence for Decision Making, 2016-2017.

Charter Member, International Prize in Statistics Foundation, 2014-2017.

Publications Officer, Statistical Learning and Data Mining Section, American Statistical Association, 2014-2016.

Chair, Ad-Hoc IMS Committee to select the editor-in-chief of the *Statistical Science*, 2012.

IMS representative to the steering committee for the International Year of Statistics, 2013.

Program Chair, Statistical Learning and Data Mining Section, American Statistical Association, 2009-2010.

Chair, ISBA Constitution and Bylaws Committee, 2009-2014

Program Chair, Institute of Mathematical Statistics, 2005.

Member, Ad-Hoc IMS Committee to select the editor of the *Annals of Applied Statistics*, 2006.

Member, Ad-Hoc ISBA Committee to select the editor of *Bayesian Analysis*, 2006.

Elected Member of the ISBA Board (International Society Bayesian Analysis), 2005-2006.

Program Chair, Statistical Computing Section, American Statistical Association, 2003-2004.

Conferences

Program Chair, ACM-IMS Conference on the Foundations of Data Science (FODS-2020).

Co-Chair, Pfizer-ASA-Columbia Symposium on Risks and Opportunities of AI in Pharmaceutical Science, May 18, 2010.

Co-Chair, ACM-IMS Interdisciplinary Summit on the Foundations of Data Science, June 15, 2019, San Francisco.

Member, NAS Leveraging Randomized Designs to Generate Real World Evidence for Regulatory Purposes Planning Committee, 2019.

Member, NAS Real World Evidence Workshop Planning Committee, 2017.

Co-Organizer, Workshop on Transdisciplinary Foundations of Data Science, Institute of Mathematics and its Applications, Minnesota, September 2016.

Chair, Organizing Committee for International Year of Statistics Capstone Workshop, 2012-13.

Member, Senior Program Committee, KDD-2011.

Member, Program Committee, 2nd International Conference on Algorithmic Decision Theory, DIMACS, 2011

Co-Chair, BioSurveillance 2007: NSF BioSurveillance Workshop: Systems and Case Studies.

Student Awards Chair, KDD-2004, *The ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Member of the Organizing Committee for the National Syndromic Surveillance Conference, 2003, 2004.

Member of the KDD-2003, *The ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Best Paper Award Committee, 2003.

Program Chair, KDD-1999, *The ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Program Chair, AISTATS-07, *Sixth International Workshop on Artificial Intelligence and Statistics*, 1997.

Member of the program committee for Association for Computing Machinery Special Interest Group on Information Retrieval Annual Conference (ACM SIGIR)-2002, 03, 04.

Member of the program committee for KDD-95, 96, 97, 98, 99, 00, 01, 02, & 2004 Knowledge Discovery in Databases.

Member of the program committee for UAI-95, 96, 97, 98, 99, 02, 03 & 04, the Annual Conference on Uncertainty in Artificial Intelligence.

Member of the program committee for AISTATS-95, 97, 99, 01, 03 & 05, the International Workshop on Artificial Intelligence and Statistics.

Member of the program committee for IDA-01, Intelligent Data Analysis.

Co-organizer, Workshop on Learning in Graphical Models, NIPS, Colorado, December, 1995.

Co-Chair for Institute of Mathematical Statistics/National Science Foundation Graphical Models Summer Workshop, 1997.

Member of the program committee for Florida Artificial Intelligence Research Symposium (FLAIRS) 1996 & 1997 Uncertain Reasoning in Artificial Intelligence Track, Florida.

Government

Chair, Scientific Advisory Committee, Insight Data Science Ireland, 2020-

Member, Organizing Committee NSF Statistics at a Crossroads, 2018.

Member, NAS Committee on Applied and Theoretical Statistics, 2017-2020.

Consultant, FDA Advisory Committees, 2014-2017.

Member, FDA Advisory Committee on Drug Safety and Risk Management, 2011-2014.

NIH Panel member, Special Emphasis Panel/Scientific Review Group 2012/10 ZRG1 PSE-B (02) M, 2012.

Member, NAS-NRC Committee on Massive Data, 2010-2011.

Member, FDA Science Board – CDER sub-committee, 2010-2011.

Science Foundation Ireland Mathematics Review Panel, March, 2005, October 2005, February 2006, November 2006, March 2007.

National Science Foundation Artificial Intelligence and Cognitive Science Review Panel, 2004

National Science Foundation Statistics Review Panel, 2003, 2004, 2007

Member of the Institute of Medicine Committee to Review the CDC Anthrax Vaccine Safety and Efficacy Research Program, 2000-2002.

Other

Member, Research Committee, National Association of Sports Officials, 2019-

Member, International and Independent External Advisory Board of the Early Detection of Neurodegenerative Diseases (EDoN) initiative, 2019-

Member, Scientific Advisory Committee, Insight Centre for Data Analytics, 2019-

External Review Committee for the School of Computer Science and Statistics, Trinity College Dublin, February 2016.

External Review Committee for the Harvard Division of Continuing Education, February 2016.

External Review Committee for Statistics (Chair), Cornell, April 2015.

External Review Committee for Arts & Sciences (Chair), Emory University, April 2015.

External Review Committee for Statistics (Chair), Carnegie Mellon University, 2013.

Member, DIMACS Advisory Board, 2013-

Member, CCICADA Advisory Board, 2012-

Member, Advisory Board, University of Maryland “Development and Evaluation of Search Technology for Discovery of Evidence in Civil Litigation.” 2011-2014.

Co-Organizer of Summer 2011 Undergraduate “Explorations in Statistics” camp, 30 students.

Member, Advisory Board, Command, Control, and Interoperability Center for Advanced Data Analysis, A Department of Homeland Security Center of Excellence, 2010-

External Review Committee for Biostatistics, Columbia University, May 2012.

External Review Committee for Statistics, Harvard University, April 2010.

External Review Committee for Statistics (Chair), Duke University, 2009.

External Review Committee for Statistics, University of North Carolina, 2008.

Editorial Board, ASA-SIAM Series on Statistics and Applied Probability, 2006-2008.

Series Editor, Chapman and Hall Computer Science and Data Analysis Series, 2002-2021.

External Examiner for PhD Dissertation of Susanne Bottcher, Aalborg University, Denmark, June 2004.

External Examiner for PhD Dissertation of Bo Thieson, Aalborg University, Denmark, September 1996.

Selected Consulting Activities

Consultant, Aris Global, 2012.

Consultant, Emergent Biosolutions, 2012-2013.

Consultant, Pharmaceutical Development Group, Inc., 2012-.

Consultant, DaVita Inc., 2010.

Consultant, Boehringer Ingelheim, 2010.

Consultant, Foundation for the NIH, OMOP, 2009-2013.

Consultant, CDC Anthrax Vaccine Research program, 2005-2010.

Consultant, Skarven Enterprises/Boeing, 2003-2012.

Consultant, Wyeth Pharmaceutical, 2006, 2008-09.

Consultant, Novartis Inc., 2006-2008.

Consultant, Takeda Inc., 2008, 2011.

Consultant, Pfizer Inc., 2007-2008, 2013-14.

Advisory Board, mediGuard/Quintiles, 2007-2011.

Consultant, Adready Inc., 2008.

Consultant, Jarvik Heart Inc., 2009-2016,

Consultant, GSK Inc., 2009.

Consultant, Eli Lilly Inc., 2016-2018.

Consultant, Endo Pharmaceuticals Inc., 2016-2017.

Consultant, Merck, 2016.

Consultant, Heron Therapeutics Inc., 2018-2019.

Consultant, Shire Plc, 2018-.

Consultant, Bayer Inc., 2019-.

Consultant, Quest Partners, 2019-.

Consultant in litigation in the last four years related to Abilify, Depakote, Ford, Incretins, Pradaxa, Talc, Taxotere, UBS, and Vioxx.

PhD Committee Chairmanships

1. Jeremy C. York, Dissertation Title: “Bayesian Methods for the Analysis of Misclassified or Incomplete Multivariate Discrete Data.” PhD Awarded 1992. Winner of the Savage Outstanding Dissertation Award. Currently employed at amazon.com, Seattle.
2. Jennifer Hoeting, Dissertation Title: “Accounting for Model Uncertainty in Linear Regression.” PhD Awarded 1994 (joint with Adrian Raftery). Currently Professor at Colorado State University.
3. Chris Volinsky, Dissertation Title: “Bayesian Model Averaging in Censored Survival Models.” PhD Awarded 1997 (joint with Adrian Raftery). Currently employed at AT&T Laboratories - Research.

4. Andrew Schaffner, Dissertation Title: “Tools for the Advancement of Undergraduate Statistics Education.” PhD Awarded 1997. Currently Professor at California Polytechnic State University, San Luis Obispo.
5. Michelle Keim, Dissertation Title: “Bayesian Information Retrieval.” PhD Awarded 1997. Currently employed at Detectent, San Diego.
6. Cibeles daSilva, Capture-recapture methodology for bowhead whales. PhD Awarded 1999.
7. Greg Ridgeway, Learning Massive Bayesian Networks. PhD Awarded 1999. Currently Associate Professor, U. Penn.
8. Susana Eyheramendy, Text categorization. PhD Awarded 2003. Currently Professor, Department of Statistics, Pontificia Universidad Catolica de Chile.
9. Ivan Zorych, Location estimation in wireless networks, PhD Awarded 2005 (NJIT/Rutgers). Previously Research Scientist at Columbia University.
10. Aimin Feng, Bayesian methods for post-marketing drug safety surveillance, PhD Awarded 2006. Currently employed at C.R. Bard, Inc.
11. Denise Chang, Individualized hospital report cards, PhD Awarded 2006. Currently employed at Sanofi-Aventis.
12. Suhrid Balakrishnan, Algorithms and Applications for Classifiers of Massive and Structured Data Problems, PhD Awarded 2007. Currently employed at Lionshare
13. Jerry Cheng, Bayesian Methods for Non-Standard Missing Data Problems, PhD Awarded 2010. Currently Assistant Professor at NYIT.
14. Shouhou Zhou, Bayesian Predictive Model Selection Criteria, PhD Awarded 2010. Assistant Professor, Penn State.
15. Shawn Simpson, Self-controlled methods for postmarketing drug safety surveillance in large-scale longitudinal data, PhD Awarded 2011. Currently data scientist at BlackRock Inc.
16. Zach Shahn, Methods for Personalized and Evidence Based Medicine, PhD Awarded 2015. Currently research scientist at IBM.
17. Ed Cheng, Applications of Bayesian Methods of Legal Problems, PhD Awarded 2018. Professor at Vanderbilt
18. Feihan Liu, PhD Awarded 2018. Data Scientist at Upstart Inc.

PhD Committee Memberships

UW PhD Reading Committees

Mayumi Adachi, Department of Music

Denise Draper, Department of Computer Science

Geof Givens, Department of Statistics

Steven Lewis, Department of Statistics

Heike Blossey, Department of Statistics

Jeremy York, Department of Statistics (Chair)

Jennifer Hoeting, Department of Statistics (Chair)

David Bradshaw, Department of Music

Craig Donovan, SIPhD
Dan Hershman, Department of Music
Tapas Kanungo, Department of Electrical Engineering

UW PhD Committees

Carlos Diaz Avalos, Department of Statistics
Lang Wu, Department of Statistics
Renato Assuncao, Department of Statistics
Dave Higdon, Department of Statistics
Brian Hopkins, Department of Mathematics
Shili Lin, Department of Statistics
Brian Lockyear, Department of Computer Science
Badr al Badr, Department of Electrical Engineering
Tapas Kanungo, Department of Electrical Engineering
Michael Heeley, Business School
Kyung-Im Sung, Department of Electrical Engineering

Rutgers PhD Committees

William McLoughlin, Department of Chemistry
Francis Mendez, Business School, RU Newark
Qi Xia, Department of Statistics
Pai-Hsi Huang, Department of Computer Science
Yihua Wu, Department of Computer Science
Aynur Dayanik, Department of Computer Science
Dmitriy Fradkin, Department of Computer Science
Ying Sun, School of Communication, Information and Library Studies
Jun Li, Department of Statistics
Yong Wang, Department of Genetics

Rutgers MS Committees

Saumitr Pathak, Department of Electrical and Computer Engineering

Columbia PhD Committees

Lucy Robinson, Department of Statistics
Jane Paik, Department of Statistics
Mladen Laudanovic, Department of Statistics
Ragna Haraldsdottir, Department of Statistics
Tyler McCormick, Department of Statistics
Kamiar Rahn timer, Department of Statistics
Pannaga Shivaswamy, Department of Computer Science
Xiaoyan Wang, Department of Biomedical Informatics
Anil Raj, Department of Applied Physics and Applied Mathematics
Ivor Cribben, Department of Statistics
Ying Liu, Department of Statistics
Maria de los Angeles Resa Juarez, Department of Statistics
Susanna Makela, Department of Statistics
David Hirschberg, Department of Statistics

Yixin Wang, Department of Statistics

Teaching

Columbia University

Research Design (APANKS5300), Fall 2017

Bayesian Data Analysis (ERM5580), Spring 2017

Applied Statistical Methods (W2025), Spring 2012, Spring 2013

Introduction to Statistical Methods (W3005), Fall 2012, Fall 2014, Fall 2015

Statistical Consulting, Fall 2008, Spring 2009, Fall 2009, Spring 2010, Fall 2010, Spring 2011

Data Mining, Fall 2008

Applied Statistics (G6101-2): Fall 2007 - Spring 2008.

Introduction to Statistical Reasoning (W1001): Spring 2010, Fall 2010.

Rutgers University

Bayesian Data Analysis: Fall 2002, Spring 2003, Spring 2004, Spring 2006

Data Mining: Spring 2001, Spring 2002, Fall 2003, Fall 2004.

Computing and Graphics in Applied Statistics (486): Fall 2001.

Mathematical Statistics (583): Spring 2003.

Mathematical Statistics (384): Spring 2007.

University of Washington

Introductory Statistics (311): Spring 1992, Fall 1992, Winter 1994, Spring 1995, Fall 1995, Spring 1997

Mathematical Statistics (341): Fall 1991, Winter 1991

Mathematical Statistics (342): Spring 1991

Introductory Statistics for Social Scientists (361-2): Fall-Winter 1998-9

Statistics for Engineers (390): Fall 1990, Spring 1993, Fall 1994, Winter 1995, Spring 1996

Stochastic Processes (396): Spring 1992

Mathematical Statistics (481): Fall 1993

Scientific Computing (535): Winter 1996, Winter 1997

Graphical Models (592): Fall 1991, Winter 1994, Spring 1995 (overload)

Statistical Consulting (598): Winter 1991, Fall 1992, Spring 1993

University Service

Columbia Service Activities

Member, Columbia University Institutional Conflict of Interest Committee , 2019-

Board Member, Columbia University Press, 2013-2016

Chair, Provost's Faculty Committee on Online Learning, 2012-2018

Chair, Columbia Shared Research Computing Policy Advisory Committee, 2011-2014

Executive Committee, Columbia Institute for Data Sciences and Engineering, 2012-

Member, Search Committee for Biostatistics Chair, 2012.

Member, Advisory Committee, Institute for Statistics and the Brain, 2012-

Member, Review Committee for Biostatistics Department, 2011.

Executive Committee, School of Continuing Education, 2011-2014.

Chair, Arts and Science Committee on Classroom Technology, 2011-2.
Member, Task Force on Benefits, 2010-2011.
Member, Provost's Committee on Retirement, 2011-2.
Member, Arts and Sciences Space Committee, 2009-2013.
Chair, Search Committee for Dean of the School of Continuing Education, 2008.
Ad-hoc promotion committees (2)

Rutgers Service Activities

Lecturer in workshop for high school teachers on Mathematics in Homeland Security, DIMACS, 2007.
Member of the Faculty of Arts and Sciences Transition Team, 2006.
Member of the Faculty of Arts and Sciences Budget Reduction Committee, 2005-06.
Member of the Faculty of Arts and Sciences Appointments and Promotions Committee, 2003-04.
Member of Organizing Committee for DIMACS Special Focus on Data Mining, 2001-.
Member of Organizing Committee for DIMACS Special Focus on Epidemiology, 2002-.
Co-Organizer of the DIMACS Working Group on Adverse Event and Disease Reporting, Surveillance, and Analysis.
Chair, DIMACS Associate Director Search Committee, 2005-06.
Member, Rutgers University Computer Coordinating Council, 2005-

UW Service Activities

Chair of the Department Computing Committee, 1994-1999.
MS Applied Exam Committee, 1991-1999 (chair in 1993)
PhD Applied Exam Committee, 1993 (reader in other years)
Graduate Program Director, 1998-1999
UW Advisory Board on Accountability, 1998-1999
Seminar Series on Graphical Models, 1990-1991 (with Russell Almond)
Multivariate Analysis and Graphical Models of Association (MAGMA) Seminar Series, 1993-1994, 1996 (with Michael Perlman)
Seminar Co-ordinator, 1991-1992 (including PNWSM)
Director of Consulting, 1993-1994
Mathday Lecture, 1994, 1995, 1997, & 1998
College of Arts and Sciences Distance Learning/Instructional Technology Task Force, 1996
College of Arts and Sciences Graduation Committee, 1996-1998.
UW Distinguished Teaching Award Committee, 1996 and 1997.
Chair, Review Committee for UW Math Science Computing, 1996.

APPENDIX 2

David Madigan

27 Colchester Street, Brookline, MA 02446

davidbmadigan@gmail.com

Tel: (862) 812-3690

Deposition and Trial Testimony – Last Four Years**Abilify**

Abilify (aripiprazole) Products Liability Litigation, DOCKET NO. BER-L-337-16, New York, June 28th, 2017 (Deposition Testimony), Pensacola, Florida, August 1st 2017 (Daubert Testimony).

Abilify Products Liability Litigation, Virtual, December 18th, 2020 (Deposition Testimony).

Ford

Ford unintended acceleration litigation, CIVIL ACTION NO.: 3:13-cv-6529 (consolidated), New York, July 26th, 2017 (Deposition Testimony).

UBS

Fernandez et al. versus UBS et al., No.: 15-cv-02859-SHS, District Court, Southern District of New York, February 5th, 2018 (Deposition Testimony), New York, August 15th 2018 (Deposition Testimony).

Pradaxa

Pradaxa litigation, Fourzon versus Boehringer Ingelheim, Superior Court of California, County of San Francisco, Case Number CGC-16-554586, New York, March 12th, 2018 (Deposition Testimony).

Taxotere

Taxotere Litigation, United States District Court, Eastern District of Louisiana, MDL No. 2740, New York, December 7th, 2018 (Deposition Testimony), September 18th, 2019 (Trial Testimony), Houston, November 14th 2019 (Deposition Testimony), Virtual, April 9th 2020 (Deposition Testimony), Virtual, August 24th 2020 (Deposition Testimony).

Incretin

Incretin Litigation, US District Court, Southern District of California, 3:13-MD-02452-AJB-MDD October 18, 2015 (Deposition Testimony), January 29, 2020 (Deposition Testimony).

Talc

Ratcliff versus Johnson & Johnson Consumer, NO. 16-2-18128-7 SEA, Superior Court of Washington for King County, New York, February 6th, 2018 (Deposition Testimony).

Ingham et al. versus Johnson & Johnson Consumer, NO. 1522-CC10417-01, Circuit Court of the City of St. Louis, New York, March 13th, 2018 (Deposition Testimony), June 20th, 2018 (Trial Testimony).

Zoas et al. versus BASF, Colgate, et al. NO. 190162/2017, Supreme Court of the State of New York, New York, July 18th, 2018 (Deposition Testimony), November 6th, 2018 (Deposition Testimony).

Leavitt versus J&J New York, November 12th, 2018 (Deposition Testimony).

Ruman and Rumondi versus BASF Catalysts Superior Court of New Jersey, MID-L-2919-17 and MID-L-2912-17, New York, January 14th, 2019 (Deposition Testimony).

Fong versus Imerys Talc Superior Court of the State of California, JCCP 4674/BC675449, New York, January 24th, 2019 (Deposition Testimony), November 18th, 2019 (Trial Testimony).

Koretzoff versus Arkema et al. New York, Superior Court of the State of California, County of Los Angeles, JCCP Case No. 4674, LASC Case No. BC 656506, January 28th, 2019 (Deposition Testimony).

Weirick. New York, Superior Court of the State of California, County of Los Angeles, April 12th, 2019 (Deposition Testimony), Long Beach, September 30th, 2019 (Trial Testimony).

Crudge, Superior Court of the State of California, County of Los Angeles, BC685901, May 29th, 2019 (Deposition Testimony), October 4th, 2019 (Trial Testimony).

Cabibi, Superior Court of the State of California, County of Los Angeles, BC665257, June 28th, 2019 (Deposition Testimony), September 9th, 2019 (Trial Testimony).

Citizen, Superior Court of the State of California, County of Los Angeles, NO. 2014-2920, 14th District Judicial Court, State of Louisiana, August 15th, 2019 (Deposition Testimony), February 4th, 2020 (Deposition Testimony).

San Nicolas, State of South Carolina for the Fifth Judicial Circuit, C/A NO.: 2017-CP-40-05764, October 1st, 2019 (Deposition Testimony), February 25th, 2020 (Deposition Testimony)

O'Hagan, Superior Court of the State of California, County of Alameda, Case No. RG19019699, October 30th, 2019 (Deposition Testimony), November 22nd, 2019 (Deposition Testimony), December 19th, 2019 (Trial Testimony).

Breakell, Superior Court Judicial District of Fairfield, County of Alameda, Case No. ASB-FBT-CV17-6066689-S, November 1st, 2019 (Deposition Testimony)

Birch, Florida, February 7th, 2020 (Deposition Testimony)

Biermann, February 20th, 2020 (Deposition Testimony)

Zimmerman, March 3rd, 2020 (Deposition Testimony), October 21st, 2020 (Deposition Testimony)

Lopez/Strain, June 19th, 2020 (Deposition Testimony)

McNeal, June 22nd, 2020 (Deposition Testimony), April 6th, 2021 (Trial Testimony)

Rayes, September 18th, 2020 (Deposition Testimony)

Bell, Lloyd et al v. American International Industries et al, September 21st, 2020 (Deposition Testimony)

Lefton, October 27^h, 2020 (Deposition Testimony)

Hamilton, January 5th, 2021 (Deposition Testimony)

Clouse, Commonwealth of Kentucky, Jefferson Circuit Court, Case No. 18-CI-03951, February 5th, 2021 (Deposition Testimony)

Chenet, Civil District Court for the Parish of Orleans, State of Louisiana, 2018-12536, February 23rd, 2021 (Deposition Testimony).

Johnson, February 26, 2021 (Deposition Testimony)

Powers, Judicial District Court for the Parish of Lafayette, State of Louisiana, 2016-0983, March 2nd, 2021 (Deposition Testimony).

Hirschberg/Eggers, March 8, 2021 (Deposition Testimony).

Metwally, March 11, 2021 (Deposition Testimony)

Foster, March 25, 2021 (Deposition Testimony)

Casaretto, April 2, 2021 (Deposition Testimony)

Prudencio, April 19, 2021 (Deposition Testimony)

Manz, June 2, 2021 (Deposition Testimony)

Pritchard, June 18, 2021 (Deposition Testimony)

Vanklive, June 25, 2021 (Deposition Testimony)

Hurley, June 29, 2021 (Deposition Testimony)

Combat Arms Earplug

3M Litigation, MDL No. 2885, Civil Action No. 3:19-md-02885-MCR-GRJ, December 11th, 2020 (Deposition Testimony).

Asbestos

George H. Wilson et al. versus Beck Arnkey Holdings LLC et al., Case No. 24x20000030, Circuit Court, Baltimore City, February 15th, 2021 (Deposition Testimony).

Zostavax

Zostavax product libailities litigation, United States District Court for the Eastern District of Pennsylvania, MDL No. 2848, Civil Action No. 2:18-md-02848, June 11th, 2021 (Deposition Testimony).

Navient

Lord Abbett Affiliated Fund, Inc. v. Navient Corp., et al., June 15th, 2021 (Deposition Testimony).